# Predicting Chemical Toxicity Effects Based on Chemical-Chemical Interactions

**Lei Chen[2]‍, Jing Lu[3]‍, Jian Zhang[4], Kai-Rui Feng[5], Ming-Yue Zheng[3]\*, Yu-Dong Cai[1]\***

**1** Institute of Systems Biology, Shanghai University, Shanghai, China, **2** College of Information Engineering, Shanghai Maritime University, Shanghai, China, **3** Drug Discovery and Design Center (DDDC), Shanghai Institute of Materia Medica, Shanghai, China, **4** Department of Ophthalmology, Shanghai First People's Hospital Affiliated to Shanghai Jiaotong University, Shanghai, China, **5** Simcyp Limited, Blades Enterprise Centre, Sheffield, United Kingdom

## Abstract

Toxicity is a major contributor to high attrition rates of new chemical entities in drug discoveries. In this study, an order-classifier was built to predict a series of toxic effects based on data concerning chemical-chemical interactions under the assumption that interactive compounds are more likely to share similar toxicity profiles. According to their interaction confidence scores, the order from the most likely toxicity to the least was obtained for each compound. Ten test groups, each of them containing one training dataset and one test dataset, were constructed from a benchmark dataset consisting of 17,233 compounds. By a Jackknife test on each of these test groups, the 1st order prediction accuracies of the training dataset and the test dataset were all approximately 79.50%, substantially higher than the rate of 25.43% achieved by random guesses. Encouraged by the promising results, we expect that our method will become a useful tool in screening out drugs with high toxicity.

## Introduction

Toxicity is a key cause of late-stage failures in drug discovery. Even some approved drugs such as Phenacetin [1] and Troglitazone [2] have been withdrawn from the market because of unexpected toxicities that were not detected during Phase III clinical trials. Thus, early toxicology data on compounds are needed to reduce R&D costs. Evaluating toxicity and assessing risks of diverse chemicals require comprehensive experimental testing against a broad spectrum of toxicity end points. These tests can cost millions of dollars, involving several thousand animals, and take many years to complete. As a result, very few chemicals have undergone the degree of testing needed to support accurate health risk assessments or meet regulatory requirements for drug approval. In recent years, the number of synthetic compounds has surged with the advance of combinatorial chemistry, and accordingly large quantities of toxicity data are urgently demanded.

Recently, particular interest has been raised to apply fast and cost-effective *in silico* toxicological models to supplement those *in vitro* and *in vivo* testing. These models require high quality toxicity data for a large set of structurally diverse drug candidates. Accelrys Toxicity is a database of toxicity information compiled from the open scientific literature [3] and containing toxicological data for approximately 0.17 million chemicals. This database is of great value for investigating the pharmacokinetic properties, metabolism and potential toxicities of compounds. Six types of toxicity data are collected in the database: (1) Acute Toxicity; (2) Mutagenicity; (3) Tumorigenicity; (4) Skin and Eye Irritation; (5) Reproductive Effects; and (6) Multiple Dose Effects. It should be noted that these categories have multiple and overlapping mechanisms of toxic action and each category represents only specific types of experiments. The combination of these experimental results may help define the overall safety profile of a compound. However, this kind of databases only provides toxicological information for recorded compounds, not for new ones. It would be valuable to accurately predict toxicities of a new compound based on the information available for recorded compounds. In order to meet the demand, there is a drive to develop quick, reliable, and non-animal-involved prediction methods, *e.g.* using structure-activity relationships (SARs) to predict drugs toxicities.

Currently, most toxicological SAR models belong to binary classifiers, which only predict compounds to be toxic or non-toxic within a single toxicity class [4,5]. It is desired to modify the strategy to predict a series of toxicity effects. In this study, we chose to build a multiclass model [6,7] to predict six categories of toxicity using the Accelrys Toxicity database instead of only one or two toxicity endpoints. However, the quadratic optimization problem

**Table 1.** Distribution of compounds in each category of compound toxicity.

| Tag | Toxicity | Total |
|-----|----------|-------|
| $T_1$ | Acute Toxicity | 12,633 |
| $T_2$ | Mutagenicity | 6,110 |
| $T_3$ | Tumorigenicity | 2,293 |
| $T_4$ | Skin and Eye Irritation | 2,353 |
| $T_5$ | Reproductive Effects | 2,501 |
| $T_6$ | Multiple Dose Effects | 4,198 |
| $T_7$ | Non-toxicity | 646 |
| Total | – | 30,734 |

doi:10.1371/journal.pone.0056517.t001

in multiclass models is difficult to solve. Thus, many previous multiclass approaches tended to decompose a multiclass problem into multiple independent binary classifications. Investigators built a set of binary classifiers, such as the model of Dietterich et al [7], each classifier distinguishing only one of the classes from the others. Although this greatly simplifies the problem, such an approach cannot provide order prediction information for the query compounds. That is, it can only predict whether the query compound has some toxicity end points, but cannot determine which is the most likely toxicity, or even the order of toxicity end points by toxicity likelihoods.

In recent years, the assessment of protein-protein interactions has been widely used to predict many attributes of proteins [8,9,10,11]. Furthermore, multiclass predictions of protein attributes have become more common [12,13,14]. These methods and their results show that interactive proteins tend to share the same functions with higher probability than do non-interactive ones.

Likewise, it is reasonable to expect that interactive compounds are also more likely to share common functions as indicated by some pioneer studies [15,16]. Thus, toxicity, as part of the biological functions of compounds, should follow the same rule. Moreover, based on a previous work on the Anatomical Therapeutic Chemical (ATC) classification of drugs [16], compared to the SAR models based on physicochemical descriptors or structural alerts, a model based on chemical-chemical interactions can rank the order of the predictions more easily and yield better prediction results. In our study, we attempt to quantify chemical-chemical interactions for each pair of interactive compounds, and obtain the confidence scores of the interactions by which the toxicity end points were ordered. Briefly, compounds of seven categories including six categories of toxicity plus non-toxicity were collected. The interactive compounds of each query compound were identified utilizing STITCH (Search tool for interactions of chemicals) [17,18]. Then, the score of each class of the query compound was obtained from the confidence scores of interactions between the query compound and its interactive compounds using the toxicity profile of the interactive compounds. Finally, the prediction quality of the model was evaluated using the Jackknife test through ten test groups. Each of these was constructed from the benchmark dataset and contained one training dataset and one external test dataset. Details are described in the following sections.

## Materials and Methods

### Benchmark Dataset

We obtained a total of 171,266 compounds from the Accelrys Toxicity Database 2011.4 [19], which had at least one toxicity effect belonging to the following six categories: (1) Acute Toxicity; (2) Mutagenicity; (3) Tumorigenicity; (4) Skin and Eye Irritation; (5) Reproductive Effects; (6) Multiple Dose Effects. Based on compound toxicity, these compounds are allocated to the 6
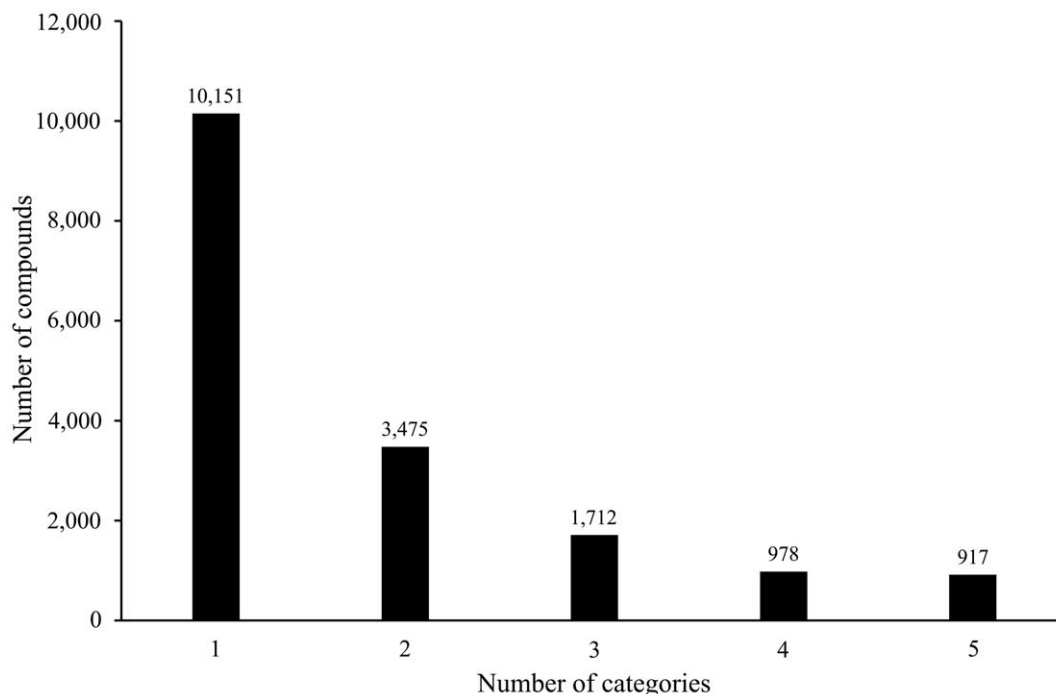


**Figure 1. The number of compounds plotted against the number of categories in the benchmark dataset.**
doi:10.1371/journal.pone.0056517.g001

categories, allowing multiple assignments. In addition, 2,871 "non-toxic" compounds including FDA-approved drugs from DrugBank [20] and endogenic metabolites from the Human Metabolome database (HMDB) [21] were collected and labeled as a negative class. For convenience, the 'non-toxic set' is regarded as the 7th category of compound toxicity. Due to lack of chemical-chemical interaction information in STITCH [17,18], some compounds cannot be investigated by this approach. After excluding these compounds, a benchmark dataset $\mathbf{S}$ consisting of 17,233 compounds was retrieved, of which 16,587 were toxic and 646 were non-toxic. These compounds are classified into 7 categories of compound toxicity. Shown in **Table 1** is the distribution of compounds in each category. The codes of 17,233 compounds and their toxicity information can be found in **Table S1**.

It is observed from **Table 1** that the sum of the number of compounds in all the 7 categories is much larger than the number of compounds, indicating that some compounds are allocated to more than one category of toxicity. Of the 17,233 compounds in the benchmark dataset, 10,151 compounds belong to only one category of toxicity, 3,475 compounds belong to two categories of toxicity, while others belong to 3–5 categories of toxicity and no compounds belong to more than five categories of toxicity - refer to **Figure 1** for a plot of the number of compounds against the number of categories of toxicity. Thus, prediction of compound toxicity is a multi-label classification problem. Like the case of processing proteins or compounds with multiple attributes [15,16,22], the proposed method would provide a series of candidate toxicities, ranging from the most to the least likely, instead of presenting only the most likely one.

To sufficiently evaluate the prediction method described in the following section, we constructed 10 test groups, denoted by $TG_1, TG_2, \ldots, TG_{10}$, respectively. In each test group $TG_i(1 \leq i \leq 10)$, there is one training dataset $\mathbf{S}_{tr}^{(i)}$ and one test dataset $\mathbf{S}_{te}^{(i)}$, i.e., $TG_i = \langle \mathbf{S}_{tr}^{(i)}, \mathbf{S}_{te}^{(i)} \rangle$, where the test dataset consisted of 1,723 compounds which were randomly selected from $\mathbf{S}$, while the training dataset contained the remaining 15,510 samples in $\mathbf{S}$, i.e., $\mathbf{S} = \mathbf{S}_{tr}^{(i)} \cup \mathbf{S}_{te}^{(i)}$ for each $1 \leq i \leq 10$. It is necessary to point out that, in each test group, the portion of the data in each class of the test dataset is roughly the same as that of the training dataset. Shown in **Table 2** is the distribution of compounds in training and test datasets of each test group.

## Chemical-chemical Interactions

It is known that two proteins that can interact with each other are more likely to share common biological functions than non-interactive ones [8,9,10,11]. Likewise, two interactive compounds are also more likely to share similar biological functions [15,16]. Since toxicity is one of a compound's properties and functions, utilizing chemical-chemical interactions to identify compound toxicity is deemed to be feasible.

The data for chemical-chemical interactions were retrieved from STITCH (chemical_chemical.links.detailed.v3.0.tsv.gz, http://stitch.embl.de/cgi/show_download_page.) [17], a well-known database including known and predicted interactions of chemicals and proteins collected from experiments, literature or other reliable sources. In the obtained file, the interaction unit contains two compounds and five kinds of scores with titles "Similarity", "Experimental", "Database", "Textmining" and "Combined_score". The last kind of score was used here to indicate the interactivity of two compounds, i.e., two compounds with "Combined_score" greater than zero were deemed interactive compounds, because the last kind of score integrates the information of the other kinds of scores. Thus, the considered

interactive compounds in this study contain the following three categories: (1) those participating in the same reactions; (2) those sharing similar structures or activities and (3) those with literature associations [17]. It is known that these categories correspond to the following three facts: (I) compounds involved in the same reactions occupy the same biological pathways; (II) compounds with similar structures or activities are likely to share similar functions, thereby occupying the same pathways with high probability; (III) the co-occurrence of two compounds, as noted in many studies, indicates some direct or indirect relationships, suggesting that they have the potential to share the same pathways. On the other hand, compounds in the same biological pathways always induce similar side effects, thereby having similar toxicity effects. Accordingly, it is reasonable to suppose that interactive compounds tend to have similar toxicity effects.

The value of the "Combined_score" of two interactive compounds indicates the likelihood that they can interact, i.e., two interactive compounds with high "Combined_score" can interact with high probability. Thus, this score is also termed a confidence score in this study. For two compounds $c_1$ and $c_2$, let us denote the confidence score of an interaction between them by $\mathbf{Q}(c_1, c_2)$. Specifically, if there is no interaction information between $c_1$ and $c_2$ based on the current records in STITCH, their interaction confidence score is assigned zero, i.e., $\mathbf{Q}(c_1, c_2) = 0$. In this study, 323,432 interaction units, i.e., 323,432 pairs of compounds with confidence scores greater than 0, were used to predict compound toxicity. The detailed information on these interaction units can be found in **Table S2**.

## Prediction Method

As is mentioned in the above section, interactive compounds are more likely to have common toxicity. Accordingly, the toxicities of a query compound can be identified according to its interactive compounds.

For convenience, let $T_1$, $T_2$, ..., $T_7$ denote the seven categories of toxicity, where $T_1$ denotes "Acute Toxicity", $T_2$ "Mutagenicity", and so forth (see column 1 and 2 of **Table 1**). Suppose that there are $n$ compounds in the training dataset, that is $c_1$, $c_2$, ..., $c_n$, the toxicity of a compound $c_i$ in the training dataset is formulated as

$$T(c_i) = [t_{i,1}, t_{i,2}, \ldots, t_{i,7}] (i = 1, 2, \ldots, n) \qquad (1)$$

where

$$t_{i,j} = \begin{cases} 1 & \text{If } c_i \text{ has toxicity } T_j \\ 0 & \text{Otherwise} \end{cases} \qquad (2)$$

Given a query compound $c_q$, its toxicity is predicted not only by its interactive compounds but also by the confidence scores of their interactions. The score indicating that the query compound $c_q$ has toxicity $T_j$ is calculated by

$$\Theta(c_q \mapsto T_j) = \sum_{i=1}^{n} \mathbf{Q}(c_i, c_q) \cdot t_{i,j}, j = 1, 2, 3, 4, 5, 6, 7 \qquad (3)$$

The high score $\Theta(c_q \mapsto T_j)$ means that there are many interactive compounds of $c_q$ in the training dataset that have toxicity $T_j$ or some interactions between $c_q$ and its interactive compounds having toxicity $T_j$ are labeled by high confidence

**Table 2.** Distribution of compounds in training and test datasets of each test group.

| | $TG_1$ | | $TG_2$ | | $TG_3$ | | $TG_4$ | | $TG_5$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Tag** | | | | | | | | | | |
| $T_1$ | 11,382 | 1,251 | 11,387 | 1,246 | 11,351 | 1,282 | 11,364 | 1,269 | 11,385 | 1,248 |
| $T_2$ | 5,475 | 635 | 5,476 | 634 | 5,529 | 581 | 5,492 | 618 | 5,491 | 619 |
| $T_3$ | 2,065 | 228 | 2,065 | 228 | 2,063 | 230 | 2,063 | 230 | 2,056 | 237 |
| $T_4$ | 2,102 | 251 | 2,102 | 251 | 2,115 | 238 | 2,112 | 241 | 2,093 | 260 |
| $T_5$ | 2,235 | 266 | 2,235 | 266 | 2,260 | 241 | 2,255 | 246 | 2,235 | 266 |
| $T_6$ | 3,747 | 451 | 3,749 | 449 | 3,777 | 421 | 3,784 | 414 | 3,799 | 399 |
| $T_7$ | 582 | 64 | 577 | 69 | 586 | 60 | 582 | 64 | 583 | 63 |
| Total | 27,588 | 3,146 | 27,591 | 3,143 | 27,681 | 3,053 | 27,652 | 3,082 | 27,642 | 3,092 |
| | $TG_6$ | | $TG_7$ | | $TG_8$ | | $TG_9$ | | $TG_{10}$ | |
| **Tag** | | | | | | | | | | |
| $T_1$ | 11,367 | 1,266 | 11,395 | 1,238 | 11,369 | 1,264 | 11,374 | 1,259 | 11,353 | 1,280 |
| $T_2$ | 5,489 | 621 | 5,500 | 610 | 5,492 | 618 | 5,497 | 613 | 5,506 | 604 |
| $T_3$ | 2,075 | 218 | 2,067 | 226 | 2,070 | 223 | 2,043 | 250 | 2,070 | 223 |
| $T_4$ | 2,123 | 230 | 2,125 | 228 | 2,135 | 218 | 2,102 | 251 | 2,133 | 220 |
| $T_5$ | 2,244 | 257 | 2,243 | 258 | 2,236 | 265 | 2,258 | 243 | 2,234 | 267 |
| $T_6$ | 3,762 | 436 | 3,750 | 448 | 3,772 | 426 | 3,777 | 421 | 3,755 | 443 |
| $T_7$ | 583 | 63 | 587 | 59 | 579 | 67 | 569 | 77 | 584 | 62 |
| Total | 27,643 | 3,091 | 27,667 | 3,067 | 27,653 | 3,081 | 27,620 | 3,114 | 27,635 | 3,099 |

scores. In view of this, the greater the score $\Theta(c_q \mapsto T_j)$, the more likely that the compound $c_q$ has toxicity $T_j$. In particular, if $\Theta(c_q \mapsto T_j)$ for some $j$, it is indicated that the probability that the query $c_q$ having the $j$-th category of toxicity is zero because there are no interactive compounds of $c_q$ in the training dataset that have toxicity $T_j$.

Since this is a multi-label classification problem, i.e., some compounds have more than one category of toxicity. A prediction method only providing the most likely toxicity is not an optimal choice. Thus, our method is valuable in that it can provide a series of candidate toxicities for a query compound, ranging from the most likely to the least likely. For example, if the results obtained from **Eq. 3** are

$$\Theta(c_q \mapsto T_3) \geq \Theta(c_q \mapsto T_1) \geq \Theta(c_q \mapsto T_6) > 0 \qquad (4)$$

it can be interpreted to mean that there are three candidate toxicities for the query compound $c_q$, and the most likely toxicity for $c_q$ is $T_3$ ("Tumorigenicity", cf. **Table 1**), followed by $T_1$ ("Acute Toxicity") and $T_6$ ("Multiple Dose Effects"). In addition, $T_3$ is called the 1st order prediction, $T_1$ the 2nd order prediction, and so forth.

### Jackknife Test

The Jackknife test [16] is often used to examine the performance of various predictors, because it can always provide a unique prediction result for a given dataset. It has been widely used by investigators to evaluate their predictors [23,24,25,26,27,28,29,30,31,32,33]. During the test, each sample in the training dataset is singled out one-by-one and tested by the predictor trained by the other samples. Thus, each sample is tested exactly once.

### Accuracy Measurement

The $j$-th order prediction accuracy is calculated by the following formula [15,16]:

$$\Gamma_j = \frac{CT_j}{N} \quad j = 1,2,3,4,5,6,7 \qquad (5)$$

where $CT_j$ denotes the number of compounds whose $j$-th order prediction is one of its true toxicities, and $N$ denotes the total number of compounds in the dataset. If a prediction method can obtain high $\Gamma_j$ with small $j$ and low $\Gamma_j$ with large $j$, it implies that the method arranges the candidate toxicities well. Among them, the 1st order prediction accuracy is the most important indicator of good or bad performance.

Although the seven prediction accuracies can be obtained by **Eq. 5**, none of them provides the overall prediction accuracy. In view of this, we employ another measurement that calculates the proportion of true toxicities of the first $m$ predictions. It can be calculated as follows [16]:

$$\Delta_m = \frac{\sum_{i=1}^{N} S_{i,m}}{\sum_{i=1}^{N} N_i} \qquad (6)$$

where $S_{i,m}$ represents the number of the correct predictions of the $i$-th compound among its first $m$ predictions, and $N_i$ represents the number of toxicities that the $i$-th compound has. Since different compounds may have different numbers of toxicities, the parameter $m$ in **Eq. 6** is usually taken as the smallest integer no less than the average number of toxicities in the dataset, which can

**Table 3.** Prediction accuracies obtained by the method as applied to training and test datasets of each test group.

| Prediction Order | $TG_1$ | | $TG_2$ | | $TG_3$ | | $TG_4$ | | $TG_5$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 79.40% | 79.69% | 79.45% | 79.28% | 79.23% | 80.62% | 79.28% | 79.45% | 79.30% | 79.34% |
| 2 | 37.16% | 38.42% | 37.14% | 38.24% | 37.54% | 37.20% | 37.17% | 38.31% | 37.40% | 36.16% |
| 3 | 22.18% | 23.16% | 22.20% | 22.87% | 22.32% | 21.65% | 22.29% | 22.63% | 22.53% | 22.87% |
| 4 | 15.45% | 16.66% | 15.49% | 16.77% | 16.35% | 14.86% | 15.46% | 16.13% | 15.41% | 15.55% |
| 5 | 11.06% | 11.61% | 11.04% | 11.49% | 11.00% | 10.85% | 10.88% | 10.16% | 10.95% | 11.20% |
| 6 | 6.92% | 7.25% | 6.84% | 7.89% | 7.23% | 5.86% | 6.99% | 6.56% | 6.85% | 7.84% |
| 7 | 1.21% | 1.33% | 1.22% | 1.04% | 1.27% | 1.51% | 1.39% | 1.45% | 1.26% | 1.68% |

| Prediction Order | $TG_6$ | | $TG_7$ | | $TG_8$ | | $TG_9$ | | $TG_{10}$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $S_{tr}^{(6)}$ | $S_{te}^{(6)}$ | $S_{tr}^{(7)}$ | $S_{te}^{(7)}$ | $S_{tr}^{(8)}$ | $S_{te}^{(8)}$ | $S_{tr}^{(9)}$ | $S_{te}^{(9)}$ | $S_{tr}^{(10)}$ | $S_{te}^{(10)}$ |
| 1 | 79.57% | 80.15% | 79.36% | 79.98% | 79.45% | 79.05% | 79.52% | 79.80% | 79.46% | 79.34% |
| 2 | 37.11% | 37.72% | 37.57% | 36.10% | 37.21% | 38.65% | 37.32% | 35.98% | 37.44% | 37.20% |
| 3 | 22.57% | 22.29% | 22.30% | 23.39% | 22.23% | 24.03% | 22.46% | 23.33% | 22.42% | 22.93% |
| 4 | 15.31% | 15.90% | 15.36% | 15.55% | 15.52% | 14.74% | 15.40% | 16.25% | 15.36% | 16.37% |
| 5 | 10.93% | 10.45% | 10.95% | 11.55% | 11.08% | 10.10% | 10.74% | 11.55% | 10.87% | 10.74% |
| 6 | 7.00% | 6.56% | 7.00% | 6.62% | 7.16% | 5.86% | 6.76% | 7.78% | 6.97% | 7.25% |
| 7 | 1.25% | 1.57% | 1.32% | 0.99% | 1.32% | 1.45% | 1.27% | 1.57% | 1.30% | 1.33% |

be computed by

$$M = \frac{\sum_{i=1}^{N} N_i}{N} \qquad (7)$$

where $m = \lceil M \rceil$. Obviously, a larger $\Delta_m$ implies better prediction performance by the method for the identification of compound toxicity.

**Table 4.** Proportions of true toxicities covered by the first two predictions for training and test datasets of each test group.

| Test group | Training dataset | Test dataset |
|---|---|---|
| $TG_1$ | 65.52% | 64.69% |
| $TG_2$ | 65.54% | 64.52% |
| $TG_3$ | 65.43% | 66.49% |
| $TG_4$ | 65.32% | 65.83% |
| $TG_5$ | 65.48% | 64.36% |
| $TG_6$ | 65.46% | 65.71% |
| $TG_7$ | 65.55% | 65.21% |
| $TG_8$ | 65.43% | 65.82% |
| $TG_9$ | 65.61% | 64.07% |
| $TG_{10}$ | 65.61% | 64.79% |

## Results

As described in the Section "Benchmark dataset", 10 test groups were constructed to evaluate the method described in Section "Prediction method". In each test group, there were one training dataset consisting of 15,510 compounds and one test dataset containing 1,723 compounds. The predicted results for each test group obtained by the proposed method are as follows.

### Performance of the Method on the Training Dataset

For the 15,510 compounds in each training dataset $S_{tr}^{(i)}(1 \leq i \leq 10)$, we conducted the prediction and evaluated its performance by the Jackknife test. Listed in the column with title $S_{tr}^{(i)}$ of **Table 3** are seven prediction accuracies, calculated by **Eq. 5**, for training dataset $S_{tr}^{(i)}$, from which we can see that the 1st order prediction accuracies were all around 79.50%, where the maximum was 79.57%, while the minimum was 79.23%; the 2nd order ones were all around 37.30%. It is indicated that the proposed method is very stable. It is also observed from the corresponding columns of **Table 3** that the accuracies followed a descending trend when increasing the order number, indicating that the method sorted the candidate toxicities quite well for the compounds in each training dataset $S_{tr}^{(i)}(1 \leq i \leq 10)$. The average
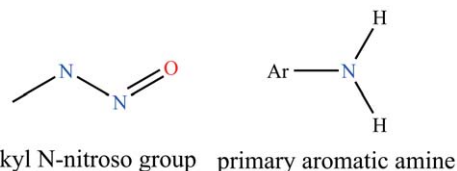


**Figure 2. The structures of the alkyl N-nitroso group and the primary aromatic amine group.**

**Table 5.** Details of Tasosartan's interactive compounds in the training dataset.

| Compound ID | Tag of toxicity class | Its interactive compound ID | Tag of toxicity class | Confidence score |
|---|---|---|---|---|
| CID000060919 | $T_7$ | CID000003749 | $T_7$ | 679 |
| CID000060919 | $T_7$ | CID000002541 | $T_7$ | 670 |
| CID000060919 | $T_7$ | CID000060921 | $T_7$ | 669 |
| CID000060919 | $T_7$ | CID000003961 | $T_7$ | 667 |
| CID000060919 | $T_7$ | CID000060846 | $T_7$ | 658 |
| CID000060919 | $T_7$ | CID000065999 | $T_1, T_6$ | 643 |
| CID000060919 | $T_7$ | CID000054738 | $T_1, T_2$ | 172 |

numbers of toxicities for compounds in each training dataset $S_{tr}^{(i)}$ were about 1.78 according to **Eq. 7**, *i.e.*, $M = 1.78$. It is noteworthy that if one predicts compound toxicity by random guesses, the average success rate would be only 25.43% (1.78/7), which is much lower than each of the 1$^{st}$ order prediction accuracies by our method. To evaluate the prediction accuracy by the method more thoroughly, **Eq. 6** was calculated by taking $m = 2$, *i.e.*, we considered the first two predictions for each compound in $S_{tr}^{(i)}(1 \leq i \leq 10)$ to see the proportions of true toxicities covered by these predictions. These proportions are shown in column 2 of **Table 4**, from which we can see that they were all about 65.50%, where the maximum was 65.61% while the minimum was 65.32%. Thus, it is indicated once again that our method is reliable.

### Performance of the Method on the Test Dataset

For the 1,723 compounds in each test dataset $S_{te}^{(i)}(1 \leq i \leq 10)$, the toxicities of these compounds were predicted by the proposed method described in Section "Prediction method" based on the compounds in the training dataset $S_{tr}^{(i)}$. After processing by **Eq. 5**, seven prediction accuracies for each test dataset $S_{te}^{(i)}$ were obtained and were listed in the column with title $S_{te}^{(i)}$ of **Table 3**. It is observed that the 1$^{st}$ order prediction accuracies were all about
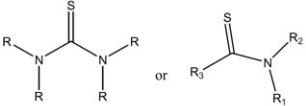
**Table 6.** The details of common compounds belonging to two categories.

| Tag of toxicity class | $T_1$ | $T_2$ | $T_3$ | $T_4$ | $T_5$ | $T_6$ |
|---|---|---|---|---|---|---|
| $T_1$ | 12,633[a] | 3,483 (22.8%)[b] | 1,485 (11.0%) | 2,027 (15.6%) | 2,075 (15.9%) | 3,446 (25.7%) |
| $T_2$ | | 6110 | 1,720 (25.7%) | 1,213 (16.7%) | 1,336 (18.4%) | 1,723 (20.1%) |
| $T_3$ | | | 2293 | 570 (14.0%) | 753 (18.6%) | 781 (13.7%) |
| $T_4$ | | | | 2353 | 731 (17.7%) | 897 (15.9%) |
| $T_5$ | | | | | 2501 | 1,409 (26.6%) |
| $T_6$ | | | | | | 4,198 |

[a]The number of common compounds belonging to two categories.
[b]The number in parenthesis means the ratio of the number of common compounds to the number of non-overlapping compounds of the two categories.

79.50%. Similar to the seven prediction accuracies for each training dataset $S_{tr}^{(i)}$, those of test dataset $S_{te}^{(i)}$ also followed a descending trend with the increase of the order number, implying that our method also arranged the candidate toxicities of samples in each test dataset quite well. According to **Eq. 7**, the average numbers of toxicities for the compounds in each test dataset were about 1.80. Thus, we still considered the first two predictions of each sample in $S_{te}^{(i)}(1 \leq i \leq 10)$ to calculate the proportions of true toxicities covered by these predictions, *i.e.*, computing **Eq. 6** by taking $m = 2$. Listed in column 3 of **Table 4** are ten proportions for ten test datasets, each yielding a probability of approximately 65%.

## Discussion

### Understanding of the Toxicity Prediction Results

It is observed from **Table 3** that the performance of the method on ten test groups is similar. Thus, the first test group (*i.e.*, $TG_1$) is used as an example to show how to interpret the toxicity predicting results in detail.

Our multiclass model achieved a quite promising performance using the chemical-chemical interactions data on test group $TG_1$ (see **Table 3** for details). For example, the compound 4-(N-methyl-N-nitrosamino)-1-(3-pyridyl)-1-butanone (CID000047289, NNK) shows positive results for five toxicity endpoints: $T_1$, $T_2$, $T_3$, $T_5$, and $T_6$. Our model accurately predicted these five kinds of endpoints, and provided the order predictions as $T_3 > T_2 > T_1 > T_6 > T_5 > T_4 > T_7$. The 7$^{th}$ label representing 'non-toxic' was ranked as the last, suggesting that this compound is very likely to have toxic effects. As stated in the Section "Chemical-chemical interactions", the interactive compounds derived from STITCH tend to have the same toxicity categories. 4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanol (CID000104856, NNAL), an interactive compound of NNK, has toxicities $T_2$ and $T_3$, which are also shared by NNK. The alkyl N-nitroso group (see **Figure 2**) of these two compounds associates with the formation of DNA adducts, and induces lung cancer in laboratory animals [34,35,36]. Another example is trimethoprim (CID000005578), which is positive for five toxicity endpoints: $T_1$, $T_2$, $T_4$, $T_5$, and $T_6$. The prediction order of our model was $T_1 > T_6 > T_2 > T_5 > T_4 > T_3 > T_7$. This compound was considered to be a carcinogen according to chemical-chemical interactions, but the Accelrys Toxicity database [19] labeled this compound only as a mutagen. However, it is reasonable to assume this compound as a carcinogen because it has a genotoxic toxicophore-aromatic amine (see **Figure 2**) [5,37,38]. Typically, mutation is one of the first steps in the development of cancer [39].
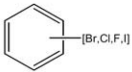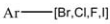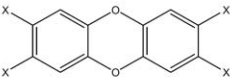
| Nongeneric SAs | Examples |
|---|---|
| thiocarbonyl | diphenylthiohydantoin (CID000854150) rubeanic acid (CID002777982) |
| halogenated benzene | 4-chlorothiophenol (CID000007815) 4-bromophenol (CID000007808) |
| halogenated PAH | 4-hydroxy-2',4',6'-trichlorobiphenyl (CID000105036) 2,2',5,5'-tetrachlorobenzidine (CID000027465) |
| halogenated dibenzodioxins | 1,2,3,4,6,7,8-heptachlorodibenzodioxin (CID000037270) 2,3,7,8-tetrabromodibenzo-4-dioxin (CID000039729) |

**Figure 3. Nongeneric SAs (Benigni) and some carcinogens matching these SAs.**
doi:10.1371/journal.pone.0056517.g003

Tasosartan (CID000060919) is an angiotensin II (AngII) receptor blocker [40], which is labeled as a relatively "non-toxic" compound in the dataset. Using our model, the order prediction of this compound was $T_7 > T_1 > T_6 > T_2$. The 1$^{st}$ order prediction is "non-toxic", consistent with the experimental data available. Among seven interactive compounds in the training dataset retrieved from STITCH (see **Table 5**), the top five interactive compounds are "non-toxic", and their confidence scores are relatively high. However, the latter two interactive compounds are toxic, so tasosartan is predicted to have some toxicity effects in our model. However, the possibility of its possessing these toxicities is less than that of its not possessing toxicity (*i.e.*, "non-toxic").

The predictions for NNK, trimethoprim, and tasosartan and the prediction accuracies of the method indicate that interactive compounds can share common toxicity with high probability, which assessment conforms to the results of predicting other attributes of compounds [15,16]. The confidence scores of chemical-chemical interactions contribute significantly to the prediction of compound toxicity. As shown in **Table 5**, the interactive compounds of tasosartan with high confidence scores dominantly have the same toxicity as tasosartan. On the other hand, the predicted results for NNK, trimethoprim, and tasosartan reflect a limitation of our model: the judgment of "toxic" or "non-toxic" is based on a collective set of compounds with interactive information. However, some compounds with low confidence scores exist and they may contribute to the input of promiscuous interaction information to the final classification model. To address this issue, a future endeavor should introduce a threshold to the interaction confidence score and exclude "noisy" information to obtain a more accurate prediction.

Moreover, many more compounds are without chemical-chemical interactions in the original Accelrys Toxicity database. It is expected that the problem of predicting compound toxicity can be solved more favorably by the method as increasing amounts of chemical-chemical interaction information become available.

## Analysis of the Relationship between Different Chemical Toxicity Effects

In the Accelrys Toxicity Database, there are 3,607 compounds with more than two types of toxicity effects and 3,475 compounds with exact two effects (refer to **Figure 1**). We analyzed the number of common compounds belonging to two categories, and the ratio of the number of common compounds to the number of non-overlapping compounds of the two categories (see **Table 6**). It can be found that the intersection of $T_5$ ("Reproductive Effects", cf. **Table 1**) and $T_6$ ("Multiple Dose Effects") is the largest, sharing 26.6% of common compounds. The overlapping compounds suggest that there may be a causal relationship between the two categories. Specifically, the reproductive effects may cause multiple dose effects, *i.e.*, reproductive toxicities may be cumulative, and hence be regarded as showing multiple dose effects in the meantime. The followed instances of correspondence between two categories are $T_2$ ("Mutagenicity") vs. $T_3$ ("Tumorigenicity") and $T_1$ ("Acute Toxicity") vs. $T_6$ ("Multiple Dose Effects"). Since, in many cases, mutation is one of the first steps in the development of cancer [39], we took $T_2$ ("Mutagenicity") vs. $T_3$ ("Tumorigenicity") as an example to study the relationship between the two toxic categories.

From the viewpoint of mechanism of action, carcinogens can be classified into genotoxic or epigenetic carcinogens. Genotoxic carcinogens can bind covalently to DNA, and many known mutagens belong to this category. In the dataset, there are 1,720 common compounds with simultaneous toxicity $T_2$ ("Mutagenicity") and $T_3$ ("Tumorigenicity"). The Structural alerts (SAs) provided by Benigni [37], which are molecular functional groups associated with a specific toxicity end point [38], were used here to gain insights into the correspondence of the two toxic effects. As summarized in **Table S3**, we illustrated a few examples for each of the matched SAs.

As previously mentioned, not all of the mutagens are carcinogens. For example, $\alpha,\beta$-unsaturated carbonyl compounds can interact with DNA by Michael addition, then lead to mutagenic and carcinogenic responses [37], *e.g.* acrylamide (CID000006579) and 2-butenal (CID000447466). However, if an

α,β-unsaturated carbonyl compound has conformational constraints or alkyl groups at the site of nucleophilic attack, the compound would be prone to reaction via Schiff base formation [41]. This change may only generate the DNA-adducts, but not undergo the following carcinogenic process [37]. This means that this kind of compound has no carcinogenicity, *e.g.* (E)-2-methyl-2-butenal (CID005321950) and 2-propylacrolein (CID000070609).

Epigenetic carcinogens do not usually bind directly to DNA, but have a large variety of different and specific mechanisms, and behave negatively in the standard mutagenicity assay [42]. Thus, some compounds that can match nongeneric SAs [37] are only carcinogens, not mutagens (see **Figure 3**).

## Conclusions

In this study, a multi-classifier for six toxicity effects was built based on 17,233 compounds with their experimental toxicity information available and 323,432 pairs of mapped chemical-chemical interaction information extracted from the STITCH database. A new chemical entity can have multiple toxicity effects, so a multiclass toxicity prediction tool may prove to be practically more valuable to chemists than a traditional binary classification model. It can provide a better toxicity profile for a compound rather than merely indicating whether the compound has a specific toxic action or potential. The outstanding performance of our approach suggests that the multi-classification scheme is feasible and effective for *in silico* chemical toxicity prediction.

## Supporting Information

**Table S1   List of 17,233 compounds investigated in this study and their toxicity information.**
(PDF)

**Table S2   List of 323,432 interaction units used to predict compound toxicity in this study.**
(PDF)

**Table S3   List of SAs (Benigni) and examples matching SAs in our dataset.**
(PDF)

## Author Contributions

Conceived and designed the experiments: LC JZ MYZ YDC. Performed the experiments: LC JL KRF. Analyzed the data: JL JZ MYZ. Contributed reagents/materials/analysis tools: LC JL KRF MYZ YDC. Wrote the paper: LC JL KRF.

## References

1. Dubach UC, Rosner B, Sturmer T (1991) An epidemiologic study of abuse of analgesic drugs. Effects of phenacetin and salicylate on mortality and cardiovascular morbidity (1968 to 1987). N Engl J Med 324: 155–160.
2. "AstraZeneca Decides to Withdraw Exanta" (2006) Available: http://www.astrazeneca.com/Media/Press-releases/Article/20060214-AstraZeneca-Decides-to-Withdraw-Exanta.Accessed 2012 Sep 2.
3. Wang WB, Zhao YP, Cong L, Jing H, Liao Q, et al. (2011) Clinical characters of gastrointestinal lesions in intestinal Behcet's disease. Chin Med Sci J 26: 168–171.
4. Zheng M, Liu Z, Xue C, Zhu W, Chen K, et al. (2006) Mutagenic probability estimation of chemical compounds by a novel molecular electrophilicity vector and support vector machine. Bioinformatics 22: 2099–2106.
5. Wang Y, Lu J, Wang F, Shen Q, Zheng M, et al. (2012) Estimation of carcinogenicity using molecular fragments tree. J Chem Inf Model 52: 1994–2003.
6. Crammer K, Singer Y (2001) On the algorithmic implementation of multiclass kernel-based vector machines. Journal of Machine Learning Research 2: 265–292.
7. Dietterich TG, Bakiri G (1995) Solving multiclass learning problems via error-correcting output codes. Journal of Artificial Intelligence Research 2: 263–286.
8. Sharan R, Ulitsky I, Shamir R (2007) Network-based prediction of protein function. Mol Syst Biol 3: 88.
9. Bogdanov P, Singh AK (2010) Molecular function prediction using neighborhood features. IEEE/ACM Trans Comput Biol Bioinform 7: 208–217.
10. Kourmpetis YA, van Dijk AD, Bink MC, van Ham RC, ter Braak CJ (2010) Bayesian Markov Random Field analysis for protein function prediction based on network data. PLoS One 5: e9293.
11. Ng KL, Ciou JS, Huang CH (2010) Prediction of protein functions based on function-function correlation relations. Comput Biol Med 40: 300–305.
12. Hu L, Huang T, Liu XJ, Cai YD (2011) Predicting protein phenotypes based on protein-protein interaction network. PLoS One 6: e17668.
13. Hu L, Huang T, Shi X, Lu WC, Cai YD, et al. (2011) Predicting functions of proteins in mouse based on weighted protein-protein interaction network and protein hybrid properties. PLoS One 6: e14556.
14. Gao P, Wang QP, Chen L, Huang T (2012) Prediction of Human Genes Regulatory Functions Based on Proteinprotein Interaction Network. Protein and Peptide Letters 19: 910–916.
15. Hu LL, Chen C, Huang T, Cai YD, Chou KC (2011) Predicting Biological Functions of Compounds Based on Chemical-Chemical Interactions. PLoS ONE 6: e29491.
16. Chen L, Zeng WM, Cai YD, Feng KY, Chou KC (2012) Predicting Anatomical Therapeutic Chemical (ATC) Classification of Drugs by Integrating Chemical-Chemical Interactions and Similarities. PLoS ONE 7: e35254.
17. Kuhn M, von Mering C, Campillos M, Jensen LJ, Bork P (2008) STITCH: interaction networks of chemicals and proteins. Nucleic Acids Res 36: D684–688.
18. Kuhn M, Szklarczyk D, Franceschini A, Campillos M, von Mering C, et al. (2010) STITCH 2: an interaction network database for small molecules and proteins. Nucleic Acids Res 38: D552–556.
19. Accelrys Toxicity Database 2011.4. Accelrys Software Inc.: San Diego, CA.
20. DrugBank. Available: http://www.drugbank.ca/downloads. Accessed 2012 Sep 2.
21. HMDB. Available: http://www.hmdb.ca/downloads. Accessed 2012 Sep 2.
22. Du P, Li T, Wang X (2011) Recent progress in predicting protein sub-subcellular locations. Expert Review of Proteomics 8: 391–404.
23. Cai YD, Lu L, Chen L, He JF (2010) Predicting subcellular location of proteins using integrated-algorithm method. Molecular Diversity 14: 551–558.
24. Shao X, Tian Y, Wu L, Wang Y, Jing L, et al. (2009) Predicting DNA-and RNA-binding proteins from sequences with kernel methods. Journal of Theoretical Biology 258: 289–293.
25. Zeng Y, Guo Y, Xiao R, Yang L, Yu L, et al. (2009) Using the augmented Chou's pseudo amino acid composition for predicting protein submitochondria locations based on auto covariance approach. Journal of Theoretical Biology 259: 366–372.
26. Chen L, Cai YD, Shi XH, Huang T (2012) Analysis of Metabolic Pathway Using Hybrid Properties. Protein and Peptide Letters 19: 99–107.
27. Esmaeili M, Mohabatkar H, Mohsenzadeh S (2010) Using the concept of Chou's pseudo amino acid composition for risk type prediction of human papillomaviruses. Journal of theoretical biology 263: 203–209.
28. Georgiou D, Karakasidis T, Nieto J, Torres A (2009) Use of fuzzy clustering technique and matrices to classify amino acids and its impact to Chou's pseudo amino acid composition. Journal of theoretical biology 257: 17–26.
29. Li BQ, Hu LL, Chen L, Feng KY, Cai YD, et al. (2012) Prediction of Protein Domain with mRMR Feature Selection and Analysis. PLoS ONE 7: e39308.
30. Jin L, Fang W, Tang H (2003) Prediction of protein structural classes by a new measure of information discrepancy. Computational Biology and Chemistry 27: 373–380.
31. Ivanciuc O (2008) Weka machine learning for predicting the phospholipidosis inducing potential. Current Topics in Medicinal Chemistry 8: 1691–1709.
32. Ravetti MG, Moscato P (2008) Identification of a 5-protein biomarker molecular signature for predicting Alzheimer's disease. PLoS ONE 3: e3111.
33. Sun XD, Huang RB (2006) Prediction of protein structural classes using support vector machines. Amino acids 30: 469–475.
34. Yuan JM, Koh WP, Murphy SE, Fan Y, Wang R, et al. (2009) Urinary levels of tobacco-specific nitrosamine metabolites in relation to lung cancer development in two prospective cohorts of cigarette smokers. Cancer Res 69: 2990–2995.
35. Kitiporn P, Jan-Phillip M, Marcus O, Fabian S, Victor S, et al. (2008) Machine learning based analyses on metabolic networks supports high-throughput knockout screens. BMC Systems Biology 2: 67.
36. Church TR, Anderson KE, Caporaso NE, Geisser MS, Le CT, et al. (2009) A prospectively measured serum biomarker for a tobacco-specific carcinogen and lung cancer in smokers. Cancer Epidemiol Biomarkers Prev 18: 260–266.
37. Benigni R, Bossa C (2011) Mechanisms of chemical carcinogenicity and mutagenicity: a review with implications for predictive toxicology. Chem Rev 111: 2507–2536.
38. Benigni R, Bossa C (2008) Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. Mutat Res 659: 248–261.
39. Arcos JC, Argus MF, editors (1995) Multifactor interaction network of carcinogenesis - a "tour guide". Boston: Birkhauser. 1–20 p.

40. DrugBank. Available: http://www.drugbank.ca/drugs/DB01349. Accessed 2012 Sep 12.

41. Patlewicz GY, Wright ZM, Basketter DA, Pease CK, Lepoittevin JP, et al. (2002) Structure-activity relationships for selected fragrance allergens. Contact Dermatitis 47: 219–226.

42. Woo YT (2003) Mechanisms of action of chemical carcinogens, and their role in structure–activity relationships (SAR) analysis and risk assessment. In: Benigni R, editor. Quantitative Structure–Activity Relationship (QSAR) Models of Mutagens and Carcinogens. Boca Raton: CRC Press. 41–80.