Drug—drug interaction through molecular structure similarity analysis

Santiago Vilar,1,2 Rave Harpaz,1 Eugenio Uriarte,2 Lourdes Santana,2 Raul Rabadan,1 Carol Friedman1

ABSTRACT

Background Drug—drug interactions (DDIs) are responsible for many serious adverse events; their detection is crucial for patient safety but is very challenging. Currently, the US Food and Drug Administration and pharmaceutical companies are showing great interest in the development of improved tools for identifying DDIs.

Methods We present a new methodology applicable on a large scale that identifies novel DDIs based on molecular structural similarity to drugs involved in established DDIs. The underlying assumption is that if drug A and drug B interact to produce a specific biological effect, then drugs similar to drug A (or drug B) are likely to interact with drug B (or drug A) to produce the same effect. DrugBank was used as a resource for collecting 9454 established DDIs. The structural similarity of all pairs of drugs in DrugBank was computed to identify DDI candidates.

Results The methodology was evaluated using as a gold standard the interactions retrieved from the initial DrugBank database. Results demonstrated an overall sensitivity of 0.68, specificity of 0.96, and precision of 0.26. Additionally, the methodology was also evaluated in an independent test using the Micromedex/Drugdex database.

Conclusion The proposed methodology is simple, efficient, allows the investigation of large numbers of drugs, and helps highlight the etiology of DDI. A database of 58,403 predicted DDIs with structural evidence is provided as an open resource for investigators seeking to analyze DDIs.

INTRODUCTION

Adverse drug events are a serious problem worldwide. In the US, they result in many injuries and deaths each year,1–3 costing millions of dollars per hospital annually and billions overall,4 5 and lead to increased hospital care.6–8 Drug–drug interactions (DDIs) are an important patient safety problem and have been reported to cause up to 50% of patient adverse events9 10 resulting in warning notices for or the withdrawal of many drugs from the market. The safety and efficacy profile of a drug can be altered significantly by the co-administration of other drugs. Multiple drug combinations in therapy are common11 and increase the risk of adverse events since concomitant drugs can share pharmacological or metabolic pathways. In extreme cases, some drugs have caused death due to the heightened adverse effect of the drug affected by the interaction. For example, cerivastatin, a drug withdrawn from the US market, caused 31 cases of fatal rhabdomyolysis prior to June 2001; the combination cerivastatin–gemfibrozil was implicated in 12 of the 31 deaths.12 Gemfibrozil causes increased blood levels of the statin resulting in a higher risk of myopathy and rhabdomyolysis.13

The development of tools to predict DDIs is important in the drug development process and in postmarketing surveillance in order to detect new drug combinations that should be contraindicated. Indeed, there is currently strong interest among regulatory authorities, such as the US Food and Drug Administration,13 and pharmaceutical companies in developing better tools for assessing drug interactions.14

The concept that similar molecules result in similar biological properties has been employed over the years by medicinal chemists.15–18 Methodologies such as QSAR/QSPR (Quantitative Structure-Activity Relationship or Quantitative Structure-Property Relationship), frequently used in computer aided drug design, are very helpful to establish relationships between the structures of molecules and their corresponding biological activity or other biological properties.16 19 Mole-cular fingerprint-based modeling has also been applied successfully to the identification of molecules structurally similar to those with a selected property.20 21 The same idea can be expanded to explain DDIs based on their structural similarities. In previous work, the similarity concept was used to develop interesting approaches comparing biological targets through the chemical similarity of their ligands.22 23

In this article we present a large-scale method for DDI discovery and prediction that uses molecular structure similarity information derived from fingerprint-based modeling. Identifying new DDIs using structural similarity is based on the basic idea that if drug A interacts with drug B, and drug C is structurally similar to A, then C should also interact with B (the argument also follows if A is replaced with B). Hence, by combining knowledge of known interactions with structural similarity it is possible to identify new interactions. As an example, it has been reported in the medical literature24 and in Micromedex25 that simvastatin, a drug that reduces levels of cholesterol by inhibiting the enzyme HMG-CoA reductase, can interact with fluconazole, a triazole antifungal drug, resulting in increased risk of myopathy or rhabdomyolysis. The methodology presented in this paper suggests new interactions by exploiting the concept that drugs similar to simvastatin can also interact with fluconazole and cause a similar effect as described above.
At the same time, drugs similar to fluconazole can interact with simvastatin causing the same mentioned effect (figure 1 illustrates this with another example). We have created a database of 58,403 new predicted interactions (not mentioned in DrugBank) for approved and experimental drugs, and have made this data resource publically available (see online supplementary tables S1–S3), which can be used by itself or in combination with other methods to identify possible candidates and improve DDI detection.

**METHODS**

**DrugBank database**

A total of 6624 drugs and 9454 DDIs mentioned in DrugBank V3.0 were used in this work.26 Drugs with more than one active ingredient, such as oxtriphylline, aminophylline, or colesevelam, and proteins and peptidic drugs were not included because molecular fingerprints are not appropriate descriptors for these types of molecules.

**DrugBank DDI database**

Drugs included in the DrugBank database were searched for possible interactions using the Interax Interaction Search engine on the DrugBank website,26 27 and duplicate DDIs from the database were eliminated. Interaction information was available for 928 drugs, resulting in a set of 9454 unique DDIs represented as follows: drug A, the description of the effect, and drug B, as shown in figure 1. The effect of the interaction associated with drug pairs was included in our analysis (eg, the DrugBank entry for the DDI tramadol–nefazodone is: increased risk of serotonin syndrome). To prepare for the calculation of DDI detection, the spreadsheet with the set of known DDIs was then transformed into a binary matrix M1 (with 928 rows and 928 columns) where a matrix cell value of 1 represented a known interaction between a pair of drugs and a value of 0 represented no interaction.

**Molecular structure similarity analysis**

Structural similarity was identified in three steps:

1. Collecting and processing drug structures: Information on the structures of the compounds in DrugBank was downloaded from the website along with the SMILE code (a chemical notation representing a chemical structure in linear textual form). The molecular structures were preprocessed using the Wash module implemented in MOE software,28 disconnecting group 1 metals in simple salts and retaining only the largest molecular fragment. The protonation state was considered neutral and explicit hydrogens were added. This step is a common process necessary to prepare the molecules for the next modeling process.

2. Structural representation: BIT_MACCS (MACCS Structural Keys Bit packed) fingerprints were calculated for all molecules included in the study.28 29 Different molecular fingerprints have been published but the basic technique is to represent a molecule as a bit vector that codes the presence or absence of structural features where each feature is assigned a specific bit position. For example, some structural features in the BIT_MACCS fingerprint for the molecule C₆H₅-C(O)-NH₂ are: bit 84 (NH₂, amine group), bit 154 (C=O, carbonyl group), bit 162 (aromatic, C₆H₅), and bit 163 (six member ring, C₆H₅).28 29

3. Similarity measures, computation, and data representation: Different measures are used to compare similarity between two molecular fingerprints. In this study, the molecular fingerprints were compared using the widely applied Tanimoto coefficient (TC).29 30 The TC can span values between 0 and 1, where 0 means ‘maximum dissimilarity’ and 1 means ‘maximum similarity.’ The TC between two fingerprint
representations A and B is defined as the number of features present in the intersection of both fingerprints A and B divided by the number of features present in the union of both fingerprints:

$$TC(A, B) = |A \cap B|/|A \cup B|$$

Using the Fingerprint Cluster module and the sim_matrix2txt.svl script in MOE, a similarity matrix $M_2$ was constructed to capture the TC measure of similarity between pairs of drugs in DrugBank (the matrix cell value represented the TC between pairs of drugs).

### Predicting new DDIs

From a technical standpoint, efficiently predicting all new DDIs reduces to matrix multiplication of the matrices $M_1$, which consists of the established interactions, and $M_2$, which consists of the similarity matrix (see step 3a in figure 2). Values in the diagonals of all the matrices are 0 because the interaction of a drug with itself is not considered. However, the same interaction can be generated at different times based on similarities obtained from different pairs, and therefore only the maximum value in the array is retained for each entry, so that the predicted interaction with the highest TC value only is considered (step 3b in figure 2). As an example, the predicted interaction voriconazole–triazolam, which increases the effect of the benzodiazepine, can be generated from the interaction voriconazole–alprazolam (the TC between triazolam and alprazolam is 0.98) or from the interaction voriconazole–midazolam (the TC between triazolam and midazolam is 0.91). In this case, the interaction associated with the highest TC value is used, and the prevailing source for the interaction voriconazole–triazolam is the interaction voriconazole–alprazolam. A symmetric transformation is carried out to obtain the final $M_3$ matrix (step 3c in figure 2), considering the highest value for each pair of drugs (note that the matrix in 3b of figure 2 is not a symmetric matrix). In the example of figure 2, interactions 1–2 and 2–3 from $M_3$ are retrieved in $M_3$ with a TC > 0.75. Interaction 1–4 is retrieved by the model with a low score (TC = 0.3). The model also predicts the new interaction 3–4 (TC = 0.9).

Once the final list of possible interactions is generated from $M_3$, the interactions are associated with the corresponding row in the initial spreadsheet containing the effect of the interaction so that the effect of the interaction can also be captured. The list of interactions predicted by the model with TC > 0.75 is given in online supplementary table S1 for the initial 928 DrugBank drugs used to construct the model. The same methodology was applied to the other drugs in DrugBank for which no interaction information was found in the Interax Interaction Search module, generating a database of new interactions for 5696 approved, nutraceutical, and experimental drugs (see online supplementary tables S2 and S3).

### Evaluation

The performance of the model was evaluated by comparing the predicted interactions based on our methodology when using different TC cut-off values with the established interactions in the initial DrugBank database. The interactions in the DrugBank database were retrieved by the method based on maximum similarity with other drug interaction pairs. The overall performance is summarized using the measures of sensitivity, specificity, precision, and enrichment factors. A receiver operating characteristic (ROC) curve has been generated for more accurate interpretation of model performance. A second evaluation by an external source other than DrugBank was also carried out for the

---

**Figure 2** Generating a drug–drug interactions (DDI) similarity model through combination of the DrugBank interaction database and molecular fingerprint-based modeling. In step 1, interaction matrix $M_1$ is created where the interactions in DrugBank are represented as “1”. In step 2, the similarity matrix $M_2$ is created based on the Tanimoto coefficient values. In step 3, $M_1 \times M_2$ is performed, the maximum value for each entry is retained, and the final matrix, $M_3$, is formed based on a symmetry-based transformation (retrieved interactions from $M_3$ when TC > 0.75 are represented in red, new predicted interactions with TC > 0.75 are represented in blue, and non-retrieved interactions from $M_3$ when TC < 0.75 are represented in green). Values in the diagonal of all the matrices are 0 because the interaction of a drug with itself is not considered.
50 most frequently sold drugs in 2009, and the performance of the method was assessed using Micromedex/Drugdex databases as a gold standard to establish the number of correct predictions.

RESULTS
Analysis of model performance using the DrugBank database
A total of 9454 DDIs were obtained from DrugBank which were associated with 928 drugs. Similarity information using molecular fingerprint-based modeling was computed for all 928 drugs and integrated into the system as described in the Methods section to develop the final model. Different cut-offs of similarity values of the TC were used to estimate sensitivity, specificity, precision, and an enrichment factor for the model. Based on a TC > 0.85, the model detected 4335 of the 9454 known interactions in the DrugBank database. It was highly unlikely that our system identified this set of 4335 interactions by chance (p < 0.0001, one-sided Fisher’s exact test). A random methodology considering the same number of possible cases (430 128 possible interactions) and the same number of true positive cases (4335) and false positive cases (6792) as predicted by our model, is capable of selecting only 245 known interactions (true positives), whereas our method identified over 17-fold more interactions. Table 1 shows the performance of our model using different cut-off values for the TC. An ROC curve containing all the possible interactions generated by the model has been plotted in figure 3 and shows an area under the curve of 0.92.

A sensitivity analysis through cross-validation was carried out by dividing the database randomly into two sets: a training set and a test set. Three evaluations were performed by moving 15%, 30%, and 45%, respectively, of the initial interactions to the test set, and by constructing the model with the remaining DrugBank interactions. Sensitivity and specificity values were calculated for the three training and test sets and showed metrics very close to the initial results using TC > 0.75 (sensitivity was 0.64, 0.61, and 0.55; specificity was 0.96, 0.97, and 0.97 for the three models, respectively; see online supplementary table S4 for more details). The robustness and the stability of the final model were barely affected by the division of two sets.

Table 1 Comparison of model performance on the DrugBank database using different cut-off values for the Tanimoto coefficient

<table>
<thead>
<tr>
<th>Tanimoto coefficient</th>
<th>TP</th>
<th>FP</th>
<th>Total (TP+FP)</th>
<th>Sensitivity</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;0.85</td>
<td>4335</td>
<td>6792</td>
<td>11127</td>
<td>0.46</td>
<td>0.39</td>
</tr>
<tr>
<td>&gt;0.80</td>
<td>5119</td>
<td>413882</td>
<td>419001</td>
<td>0.98</td>
<td>17.73</td>
</tr>
<tr>
<td>&gt;0.75</td>
<td>3894</td>
<td>410656</td>
<td>414550</td>
<td>0.98</td>
<td>16.24</td>
</tr>
<tr>
<td>&gt;0.70</td>
<td>6442</td>
<td>18366</td>
<td>24808</td>
<td>0.68</td>
<td>0.26</td>
</tr>
<tr>
<td>&gt;0.60</td>
<td>3012</td>
<td>402308</td>
<td>405320</td>
<td>0.96</td>
<td>11.81</td>
</tr>
</tbody>
</table>

FN, false negatives; FP, false positives; Precision, TP/TP+FP; Sensitivity, TP/TP+FN; Specificity, TN/TN+FP; TN, true negatives; TP, true positives.

Prediction of the effect produced by the DDI
Another feature of the model is its ability to detect the biological effect produced by the DDI. As an example, an interaction could produce an effect based on alterations in the bioavailability of one of the drugs due to both drugs being metabolized by the same enzymes or due to competition for the same transporter protein. In order to verify whether the model is also capable of predicting the effect produced by the DDI, a random selection of DrugBank interactions was reviewed manually to determine the degree of precision of the predicted biological effect. Out of 100 interactions selected using a TC cut-off value of 0.85, the effect produced by the drug combination was correctly predicted in 99 interactions where the effect was the same as that originally specified in DrugBank. Using other cut-offs, that is, 0.85 ≥ TC > 0.80 and 0.80 ≥ TC > 0.75, the model correctly predicted the effect in 96% and 91% of the evaluated interactions, respectively (see online supplementary table S5 for more details). However, in future predictions the nature of the predicted interactions should be carefully analyzed, especially when the TC is lower and the pharmacological class of the drugs detected as structurally similar is different. For this reason, for values of TC < 0.85, appropriate pharmacological knowledge to correctly interpret the effect of the interaction predicted would be beneficial.

Evaluation in Micromedex/Drugdex
In the second part of the evaluation, interactions for the 50 most frequent commercial drugs (consisting of 44 unique generic drugs) sold in 2009 were searched in the Micromedex/Drugdex database. Table 2 provides details of the sources of the drug information as well as the results (see also figure 4). Specifically, table 2A gives the number of interactions specified in DrugBank and in Micromedex/Drugdex; table 2B gives the number of predicted interactions and the number of interactions correctly predicted by our model; and table 2C gives the sensitivity, specificity, precision, and enrichment factor for the three different TC cut-off values. A total of 1760 interactions were associated with the drugs specified in Micromedex, and the model predicted 548 interactions with a TC > 0.75 (31% correct classification) and 348 interactions with a TC > 0.85 (20% correct classification). Detailed results are given in table 2 and online supplementary tables S6 and S7. It was highly unlikely that our
Research and applications

Table 2: Model performance predicting the interactions for the 50 most frequently sold drugs in 2009 based on the Micromedex/Drugdex database as a gold standard and using different cut-off values for the Tanimoto coefficient

<table>
<thead>
<tr>
<th>Model performance (Tanimoto coefficient &gt;0.85)</th>
<th>TP</th>
<th>FP</th>
<th>Total (TP+FP)</th>
<th>Sensitivity</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>348</td>
<td>793</td>
<td>1141</td>
<td>0.20</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>FN</td>
<td>TN</td>
<td>Total (FN+TN)</td>
<td>Specificity</td>
<td>Enrichment factor</td>
<td></td>
</tr>
<tr>
<td>1412</td>
<td>6137</td>
<td>62791</td>
<td>0.99</td>
<td>11.08</td>
<td></td>
</tr>
</tbody>
</table>

A
- Interactions described in DrugBank: 773
- Interactions described in Micromedex/Drugdex: 2323
- Interactions described in Micromedex with DrugBank drugs: 1760

B
- Tanimoto coefficient >0.85
- Interactions predicted by the model:
  - Number of correct interactions predicted by the model: 1141
  - Number of correct interactions predicted by the model and described in DrugBank: 212
  - Number of correct interactions predicted by the model and not described in DrugBank (ie, new for DrugBank): 136
- Interactions not described previously in DrugBank and correctly predicted by the model: 1141
- Sensitivity: 0.20
- Enrichment factor: 11.08

C
- Model performance (Tanimoto coefficient >0.80): 444
- Total (FP+FN): 1591
- Sensitivity: 0.25
- specificity: 0.28
- Enrichment factor: 11.04
- Model performance (Tanimoto coefficient >0.75): 1316
- Total (FP+FN): 62341
- Sensitivity: 0.98
- Specificity: 10.14
- Enrichment factor: 7.65

*The evaluation is carried out using the interactions described in Micromedex/Drugdex when the drugs implicated in the interaction are also present in DrugBank and the molecular structure is also available so the likelihood of the interaction can be computed.

†Predicted interactions using the 1454 drugs from DrugBank (approved and nutraceuticals).

FN, false negatives; FP, false positives; Precision, TP/(TP+FP); Sensitivity, TP/(TP+FN); Specificity, TN/(TN+FP); TN, true negatives; TP, true positives.

model identified 348 true interactions by chance (p<0.0001, one-sided Fisher’s exact test). A random method considering 63,932 possible interactions (interactions generated between 1454 drugs from DrugBank and the 44 most frequently sold drugs in 2009) and randomly selecting 1141 positive cases (the same as the model when TC>0.85) would detect 31 interactions described in the Micromedex database (1.78% correct classification).

The results identify interesting drug interactions belonging to two categories. The nature of the system permits the identification of drugs belonging to pharmacological classes different from those of the drugs implicated in the interaction (eg, drug A and a similar drug C do not belong to the same pharmacological class but each interact with drug B), which occurs more frequently as the TC value decreases. However, the method is more likely to identify interactions between drugs with similar pharmacological profiles. The information provided by the model in this case is more obvious but still could be very useful to researchers, particularly those without a strong background in pharmacology.

The interaction examples shown below were predicted by our model and not described in DrugBank, but were described in Micromedex/Drugdex with different levels of documentation,

from ‘the existence of the interaction was clearly established through controlled studies’ to ‘limited documentation but pharmacological knowledge lead clinicians to recognize the possible interaction’.

Examples of different pharmacological classes

Several interaction examples predicted for the 50 most frequently sold drugs in 2009 showed that the DDI similarity model can detect drugs that belong to different pharmacological classes but have similar structural features (see table 3). An example of an interaction correctly predicted by the model according to the Micromedex/Drugdex database is aripiprazole—mometasone. Concomitant use of these drugs can cause increased concentration of aripiprazole. Our model detected this interaction because the interaction aripiprazole—itraconazole is described in DrugBank, where the result is that itraconazole increases the effect of aripiprazole. According to our analysis, itraconazole shows some structural features similar to nefazodone (TC=0.82), although both molecules have different pharmacological profiles (itraconazole is an antifungal and nefazodone is an antidepressant).

Another example of an interaction found by our methodology is mometasone and different protease inhibitors used in HIV therapy, such as indinavir, neflavin, ritonavir, and saquinavir, possibly increasing the effect and toxicity of mometasone (see table 3). Mometasone is similar to fusidic acid (TC=0.77), and it is established in DrugBank that fusidic acid can interact with protease inhibitors. The possible interaction mometasone—protease inhibitors is described in Micromedex/Drugdex and may cause increased mometasone plasma concentrations due to inhibition of CYP3A4-mediated mometasone metabolism by the antiretroviral drugs.

Buprenorphine, an opioid analgesic, has been found to share some structural similarity with vinblastine, an antineoplastic agent used for the treatment of different types of cancer.
The model correctly predicts, based on Micromedex/Drugdex, that buprenorphine can interact with different protease inhibitors (atazanavir, darunavir, indinavir, ritonavir, and saquinavir), with the antifungal ketoconazole and with the macrolide antibiotic erythromycin, causing decreased metabolism of buprenorphine and increased drug plasma concentrations (see table 3).

Different predictions predicted by our model and described in Micromedex/Drugdex have been found for venlafaxine, an antidepressant of the serotonin-norepinephrine reuptake inhibitor (SNRI) class. According to our fingerprint-based model, tramadol was found to be similar to venlafaxine with TC=0.93. Therefore, venlafaxine was predicted to interact with cimetidine, clozapine, haloperidol, and dextroamphetamine, producing different plasma concentrations of the drugs implicated in the interaction (see table 3).

The possibility of finding drugs belonging to different classes increases as the TC value decreases, which is interesting. However, since the similarity is lower, the risk of incorrect predictions is higher. For this reason, we considered a cut-off value of 0.75 for the TC appropriate since similarity is still remarkable and many different classes of related drugs can be identified.

Examples of the same pharmacological classes
Although the DDI model can associate drugs which have different pharmacological profiles but are structurally similar, some of the predicted interactions can identify a drug belonging to the same pharmacological class of one of the drugs implicated in the known interaction. The DrugBank database describes the interactions acetophenazine—cisapride and acetophenazine—terfenadine as resulting in an increased risk of cardiotoxicity and arrhythmias. Our model detects that acetophenazine, a first generation antipsychotic of the phenothiazine class, is similar to quetiapine, a second generation antipsychotic, with TC=0.78.

<table>
<thead>
<tr>
<th>Drug A</th>
<th>Interaction Drug A-Drug B</th>
<th>Drug B</th>
<th>Drugs predicted to interact with Drug B because they are similar to Drug A</th>
<th>TC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Itraconazole</td>
<td>Drug A can cause increased drug B concentration</td>
<td>Aripiprazole</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Fusidic Acid</td>
<td>Drug B can cause decreased drug A metabolism (possible increased effect and toxicity of drug A)</td>
<td>Indinavir Nelfinavir Ritonavir Saquinavir</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Vinblastine</td>
<td>Drug B can cause decreased drug A metabolism</td>
<td>Atazanavir Darunavir Indinavir Ritonavir Saquinavir Erythromycin Ketoconazole</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>Buprenorphine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3 Examples of new predictions for the top 50 drugs not described in DrugBank but detected by our method and confirmed through Micromedex/Drugdex
Quetiapine is predicted to have the same interactions, which were confirmed in Micromedex/Drugdex. Other examples of predictions validated in Micromedex are the reduction in hydrochlorothiazide absorption due to concomitant use of colestipol, fenofibrate may increase the anticoagulant effect of phenprocoumon with risk of excessive bleeding, and buprenorphine can interact with different opioids resulting in precipitation of withdrawal symptoms (see table 3 and online supplementary tables for more details).

### DISCUSSION

Different types of models for predicting DDIs have been recently published. However, the majority of the in silico approaches to predicting drug interactions have focused on the integration of in vitro data to generate models for the in vivo prediction of drug interactions. These models mainly try to predict possible metabolic interactions, especially interactions related to CYP enzymes. Nevertheless, there are many examples of drugs that follow other metabolic routes. There are also many

<table>
<thead>
<tr>
<th>Drug B can cause decreased drug A metabolism and clearance</th>
<th>Cimetidine Clozapine Dextroamphetamine Haloperidol</th>
<th>Venlafaxine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tramadol</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Acetophenazine</strong></td>
<td>Increased risk of cardiotoxicity and arrhythmias</td>
<td><strong>Cisapride</strong></td>
</tr>
<tr>
<td><strong>Trichlormetiazide</strong></td>
<td>Reduced absorption of drug A</td>
<td><strong>Colestipol</strong></td>
</tr>
<tr>
<td><strong>Warfarin</strong></td>
<td>Drug B may increase the anticoagulant effect of drug A. Risk of bleeding</td>
<td><strong>Fenofibrate</strong></td>
</tr>
<tr>
<td><strong>Naltrexone</strong></td>
<td>Precipitation of withdrawal symptoms</td>
<td><strong>Alfentanil</strong></td>
</tr>
</tbody>
</table>

TC is the Tanimoto coefficient that measures similarity between two drugs.
DDIs due to similar distribution profiles of the investigated drugs. The importance of some mechanisms, as interaction with transporters, has been recognized later. 14

We propose a large-scale method based on identifying molecular similarity to analyze multiple types of drug interactions caused by the inhibition of metabolizing enzymes, transporters, or even the pharmacological targets. The model described in this article can exploit experimental knowledge to identify the possible causes of the interaction. The system allows the researcher to monitor the data and the model’s predictions preserve the nature of the original DDI that generates the outcome, which is very useful for examining the effect and the type of interaction predicted. Indeed, we reviewed 300 randomly selected interactions and have shown that the system can predict the effect of the interactions in more than 90% of cases when TC > 0.75 (see online supplementary table S5).

The model identifies a visible pattern in the DrugBank database (similar drugs have similar interactions) by detecting drugs similar to the drugs implicated in the interactions described previously. Therefore, one limitation of this study is that the performance of the model depends on the comprehensiveness of the information in the original interaction database. This method was applied to the interactions and drugs only specified in DrugBank, but the addition of other sources of established DDIs, such as those mentioned in drug labels, could be taken into account to generate the final model.

An additional issue is that 2D similarity fingerprints were used, which have some limitations in describing the molecular structure. The 3D structure is a very important component in the interaction drug–receptor and is a better representation of the molecules. 15 16 However, although the information provided by 2D methods is more limited than the 3D information, the 2D methods still offer good results and are much simpler and require less computational effort, avoiding important problems such as the selection of bioactive conformations and the calculation and superimposition of the 3D structure of all the drugs implicated in the study. Different 2D molecular fingerprints could also be used in the development of this type of model. 17 Nevertheless, in the current study, BIT_MACCS fingerprints were calculated because they are simple and have offered good results for recognizing similar molecules in large databases. 21 28 29

Although the similarity model provides valuable information associated with the initial interactions, a more reliable and complex system could be implemented through the integration of structural similarity measures and knowledge in pharmacological databases containing information about possible targets and metabolizing or transporter enzymes. This method could also be combined with other methodologies using different types of information, such as the Food and Drug Administration’s Adverse Event Reporting System, 40 which was created to provide postmarketing drug safety information, or the use of clinical data in electronic health records. 41 An extensive database of annotated possible drug interactions predicted by our model for the drugs in DrugBank (approved and experimental drugs) is provided in online supplementary tables S1–S3). This database is a valuable source of information on drug interactions that is available for download and can be used by itself or in combination with other methods to filter out possible candidates and improve DDI detection.

Several DDIs highlighted by our methodology were not known and consequently were considered false positives in our evaluation. However, it is possible that some of these drugs actually do interact but have not yet been identified. Therefore, it is possible that the false positive rate is lower than we estimated.

CONCLUSION
The results presented in this study demonstrate the usefulness of the proposed drug–drug interaction methodology as a promising approach for in silico prediction of drug interactions and their effects. The method described in this article is very simple, efficient, applicable to large-scale investigation and helps highlight the etiology of DDIs (see table 3). In this study, the application of structure similarity information to drug interaction knowledge as specified in DrugBank led to retrieval of the majority of known interactions, showing a sensitivity of 0.68 when the specificity was 0.96. A set of interactions not described in the literature but with strong supporting evidence according to our model has been constructed for further analysis. Experimental drugs were also evaluated by the model and ranked according to interaction probability. The database of SS-403 new predicted DDIs provided in this study could be useful for further study of possible candidates, and is available for download (online supplementary tables S1–S5). This database could be used as a powerful pharmacovigilance tool by itself or combined with other methods, such as the Food and Drug Administration’s Adverse Event Reporting System or electronic health records, to facilitate drug safety by selecting candidates with a strong possibility of interacting in the human body.

Contributors SV and CF conceived and designed the study; SV, RH, EU, LS, RR, and CF provided data and analysis tools; SV performed and analyzed the data; and SV, RH, EU, LS, RR, and CF wrote the paper.

Funding This work was supported by grants R01 LM010016 (CF), R01 LM010016-051 (CF), R01 LM010016-052 (ICF), R01 LM008625 (CF), and 1R01LM010140-01 (RR) from the National Library of Medicine, Plan Galego de Investigación, Innovación e Crecemento 2011-2015 (ICF), the European Social Fund (ESF), and the Angelés Alvarín program from Xunta de Galicia (Spain).

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES


1073
Research and applications

29. MACCS Keys; MDL Information Systems I, 14600 Catalina Street, San Leandro, CA 94577.
36. MCLM Keys; MDL Information Systems I, 14600 Catalina Street, San Leandro, CA 94577.
40. MACCS Keys; MDL Information Systems I, 14600 Catalina Street, San Leandro, CA 94577.