Molecular Docking for Prediction and Interpretation of Adverse Drug Reactions

Heng Luo, Achille Fokoue-Nkoutche, Nalini Singh, Lun Yang, Jianying Hu, Ping Zhang

Center for Computational Health, IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA; Cognitive Computing, IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA; Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; Bio-X Institutes, Shanghai Jiao Tong University, Shanghai 200030, China

Abstract: Adverse drug reactions (ADRs) present a major burden for patients and the healthcare industry. Various computational methods have been developed to predict ADRs for drug molecules. However, many of these methods require experimental or surveillance data and cannot be used when only structural information is available. We collected 1,231 small molecule drugs and 600 human proteins and utilized molecular docking to generate binding features among them. We developed machine learning models that use these docking features to make predictions for 1,533 ADRs. These models obtain an overall area under the receiver operating characteristic curve (AUROC) of 0.843 and an overall area under the precision-recall curve (AUPR) of 0.395, outperforming seven structural fingerprint-based prediction models. Using the method, we predicted skin striae for fluticasone propionate, dermatitis acneiform for mometasone, and decreased libido for irinotecan, as demonstrations. Furthermore, we analyzed the top binding proteins associated with some of the ADRs, which can help to understand and/or generate hypotheses for underlying mechanisms of ADRs.

Keywords: molecular docking, chemical-protein interactome, machine learning, prediction, side effects, adverse drug reactions

1. INTRODUCTION

Adverse drug reactions (ADRs) are injuries (side effects) resulting from normal doses of medications. ADRs present a major public health problem, leading to 700,000 emergency room visits and 120,000 hospitalized patients per year [1]. ADRs also cause withdrawal or restricted use of drugs, impacting both the revenue of the pharmaceutical industry and drug availability for patients.

During drug development, candidate drugs are typically tested on animal models and small human cohorts to be assessed for toxicities and safety profiles [2]–[5]. However, after passing all preclinical studies and clinical trials, drugs still commonly cause ADRs when they enter the market; the US Food and Drug Administration Adverse Event Reporting System (FAERS) has received an increasing number of ADR reports each year since 1968 [6]. ADRs are hard to predict because the mechanisms by which they occur are complicated and vary based on patient genetics [7]–[14]. Though ADR prediction is challenging, researchers have explored a variety of approaches to predict ADRs, including structure-based prediction [15]–[19], gene expression analysis [20], networks [21] and data mining of post-market surveillance data, electronic health records and social media [22]–[24]. While most of these methods require the drug molecules to be tested in experiments or on the market in order to generate sufficient data for ADR prediction, structure-based methods only require knowledge of the drug conformation. Thus, structure-based prediction has the advantage of identifying ADRs early during drug development.

With recent increases in computing power, machine learning has played an important role in healthcare data analysis and prediction. In previous studies, we utilized machine learning models to predict drug indications [25] and drug-drug interactions [17] based on the drug-target binding profiles generated from molecular docking. We showed that the chemical-protein interactome (CPI), a collection of drug-target binding profiles simulated with molecular docking, can be used to both predict drug pharmacologic actions and to provide possible biologic explanations for them. In this study, we explore whether we can apply techniques from machine learning to molecular docking scores for ADR prediction. Further, we use this approach to generate hypotheses to better understand the mechanisms of ADRs.

Previous work has used either machine learning or molecular docking to study ADRs. Luo et al. implemented molecular docking to explore associations between drug molecules and human leukocyte antigens for ADR prediction [18]. However, in the absence of a machine learning model, ADRs could not be directly linked to or predicted from drug molecules. In contrast, Liu et al. performed large-scale prediction of ADRs using features including chemical fingerprints, biological targets, enzymes, indications and other phenotypic properties [26]. However, this method cannot be used for drug candidates early in the development pipeline, for which biological and phenotypical features are not available. In our study, we combine molecular docking analysis with techniques from machine learning to predict

*Corresponding author: Ping Zhang, IBM T.J. Watson Research Center, 1101 Kitchawan Rd, Yorktown Heights, NY 10598, USA; Tel: (914) 945-4498; Email: pzhang@us.ibm.com

© 2014 Bentham Science Publishers
and generate possible explanations for ADRs using only structural information. Compared to previous work [27], we analyze a larger protein collection and further generate explanations of prediction results using docking score rankings and statistical analysis. We demonstrate that our method is able to both predict ADRs using only molecular structures and provide potential biological explanations for these ADRs.

2. MATERIALS AND METHODS

2.1. Preparation and molecular docking of drug and protein sets

The overall workflow of this study is shown in Figure 1. First, we prepared drug molecules and protein structures for molecular docking. For the drug set, we harvested the SMILES code of all small molecules in DrugBank 4.3 [28] and generated their 3D structures using Molconvert in Marvin Beans 16.10.3. We removed drug molecules that did not have rotatable bonds (e.g., calcium acetate) or were too large (molecular weight > 1200 g/mol, e.g., cisatracurium besylate), as these molecules are too large to fit into protein pockets and thus do not generate meaningful docking scores. For the protein set, we harvested all human proteins within the general collection of proteins from the PDBBind 2015 database [29], a curated source of crystal structures. For each protein, we picked the unique structure with the best available resolution.

After selection of the drug and protein sets, we prepared structure files and added Gasteiger charges using AutoDock Tools 1.5.6 [30]. The binding pockets of the proteins were centered at the original embedded ligands from the crystal structures with a fixed size of 25x25x25 Å to reduce pocket-based variation. Each of the drug molecules was docked to each of the protein structures using AutoDock Vina 1.1.2 [31] with a fixed random seed and default parameters. The lowest docking scores and corresponding binding conformations were extracted from the docking results. This process yielded a feature matrix that contains drugs as rows, proteins as columns, and the binding scores as features.

2.2. Development and evaluation of machine learning models

We harvested data from the SIDER 4.1 database [32], which contains ADR information extracted from drug labels, as our ground truth for ADR labels. The drug names from SIDER were mapped to DrugBank IDs using DrugBank synonyms. If a drug was known to cause a certain ADR, that drug-ADR pair was marked as “1” (positive); otherwise, it was marked as “0” (negative). Thus, we harvested a binary label matrix that contained drugs as rows and ADRs as columns. Each element in the matrix contained either a “1” (positive) or “0” (negative). Then, we removed any ADRs matched with less than five positive drugs, as these ADRs did not have enough positive samples to effectively learn a classifier. We then developed one logistic regression classifier for each ADR, using the protein binding scores as features. The classifiers were implemented in Python 2.7.12 (Anaconda 4.1.1) with sklearn 0.17.1. We explored different combinations of regularization types (L1 and L2) and regularization parameters (C = 0.001, 0.01, 0.1, 1, 10, 100 and 1000) during 10-fold cross-validations to avoid model overfitting. To evaluate the model performance, we used both the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPR). Both values compare the prediction values against the label values (truth) for performance evaluation. The receiver operating characteristic (ROC) curve plots the true positive rates against the false positive rates of classifiers that discriminate between positive and negative examples at various thresholds. The area under the ROC curve, or AUROC, evaluates how well the predictive model performs on the data; a random predictor would obtain an AUROC value of 0.5 while a perfect model would obtain an AUROC value of 1.0. The precision-recall curve is a plot of classifier precision as a function of classifier recall, again computed for various thresholds of the classifier. A larger area under the precision-recall curve (AUPR) value indicates a better model. In our experiment, we selected the best regularization parameters based on the largest AUROC value, and provided AUPR values as references.

Structural fingerprints are encoding systems that convert molecular structures into a series of numbers to represent their structural features. Like molecular docking scores, computing structural fingerprints requires no more information than the structures. We generated seven different types of structural fingerprints used in our previous work [25] to serve as baselines for comparison of our molecular docking model. The seven structural fingerprints are E-state, Extended Connectivity Fingerprints (ECFP)-6, Functional-Class Fingerprints (FCFP)-6, Fingerprint 4 (FP4), Klekota-Roth method, Molecular ACCess System (MACCS) and PubChem structural descriptors (labeled E-state, ECFP6, FCFP6, FP4, KR, MACCS and PubChem, respectively). We compared the prediction performance of molecular docking features against these structural fingerprints via 10-fold cross-validations using both AUROC and AUPR evaluations.

After this comparison, we used the best molecular docking models to generate ADR predictions for drugs that did not exist in our training set. To determine which protein binding features were contributing to these ADR predictions, we ranked the docking scores by their values. Further, we conducted Analysis of Variance (ANOVA) tests on the docking scores and their corresponding labels. For each ADR, scores of all the drugs corresponding to each protein feature were grouped by label (positives vs negatives) and compared against each other using ANOVA. If a protein feature is important for an ADR, the docking scores corresponding to positive labels should be significantly lower (or more energetically favorable) than those corresponding to negative labels. We calculated means of the two groups and p values using ANOVA in order to rank protein features to better understand their contributions towards a given ADR.

3. RESULTS and DISCUSSIONS

3.1. Machine learning model for ADR prediction based on SIDER

We harvested 1,231 small-molecule drugs from DrugBank and 600 unique human proteins from PDBBind (Supplementary Tables S1 and S2), generating a 1,231 by
600 docking score matrix. After mapping and filtering, the SIDER database contained 655 drugs that exist in our docking score matrix and 1,533 ADRs that have at least 5 or more positive drugs (Supplementary Table S3). We used the 655 drugs present in both datasets and 600 protein binding features to develop 1,533 logistic regression classifiers for the corresponding 1,533 ADRs (Supplementary Tables S1-S3). During 10-fold cross-validation, the models based on molecular docking obtained overall AUROC = 0.843, AUPR = 0.395, accuracy = 0.906 and F1 score = 0.431 with the optimal L2-regularization parameters based on all drug-ADR pairs (micro-averaging). As a comparison, the molecular docking-based models outperformed all the structural fingerprint-based models in terms of both AUROC and AUPR (Figure 2), demonstrating reasonably good overall performance for the ADR prediction task.

For each ADR, we evaluated the individual performances of the machine learning models and obtained the top 10 best performing models ranked by the AUROC values in Table 1. We found that each of the top 10 prediction models obtained high AUROC values (> 0.95), and most of them also had good F1 scores (> 0.65). Based on these AUROC values and F1 scores, we believe our docking features are useful for predicting these top ranking ADRs.

3.2. Case studies of the prediction results

For 576 drugs that have docking score features but do not exist in our training set, we used our machine learning models to make predictions for the 1,533 ADRs, generating 576 by 1,533 prediction values. By looking at the top resulting predictions, we believe that we can discover some potential ADRs. Next, we discuss four case studies from these predictions that show how our models can be used to predict and explain these ADRs.

3.2.1. Prediction of fluticasone propionate-induced skin striae

Fluticasone propionate (DrugBank ID: DB00588), a glucocorticoid for skin disease, was predicted by our model to cause skin striae (UMLS Concept ID: C0152459) with a high confidence score of 0.811. This confidence score ranked top 1 among the drugs (row-wise), as well as top 2 among the ADRs (column-wise). The ADR prediction model for skin striae obtained an AUROC = 0.884, AUPR = 0.446, accuracy = 0.980 and F1 score = 0.667, each of which indicate reasonably good prediction performance. This predicted result is supported by the literature, which has shown that topical glucocorticoids can cause adverse effects such as skin atrophy, skin striae, and telangiectasia [33].

3.2.2. Prediction and possible interpretation of mometasone-induced dermatitis acniform

Dermatitis acniform (UMLS Concept ID: C0234708) is a term describing acne-like cutaneous eruptions. Our prediction results showed that mometasone (DrugBank ID: DB00764) was the highest-ranked drug in our test set to cause this ADR, with 0.649 confidence. It has been previously reported that acneiform eruption is a local adverse effect caused by mometasone use [34], validating our prediction. To understand the potential mechanisms of this ADR, we looked at the top binding proteins for mometasone ranked by binding scores (Table 2). The orphan nuclear receptor gamma (RORγt) ligand-binding domain (Protein Data Bank ID, or PDB ID: 3B0W) was predicted to be the top binding target for mometasone with a binding score of -10.4 (Figure 3a). Studies from the literature have found that IL-17 expressing cells and Th17-related signaling exist in acneiform lesions [35], and that RORγt is needed for Th17 cell differentiation and IL-17 production [36], [37]. Thus, it is possible that mometasone induces the occurrence of dermatitis acniform by binding to RORγt and affecting the Th17/IL-17 level.

3.2.3. Prediction and potential explanation of irinotecan-induced decreased libido

We found that a colorectal cancer drug, irinotecan (DrugBank ID: DB00762), was predicted as a top drug to cause decreased libido (UMLS Concept ID: C0013922) with a 0.823 confidence. The literature suggests that this is a valid result, as studies have found that decreased sex drive is a common side effect of cancer drugs [38]. We looked at the top binding targets for irinotecan (Table 3) and found that a cancer-related target, tumor necrosis factor (TNF) family cytokine CD40 ligand (CD40L), was ranked to the top 1. It has been reported that CD40L can significantly enhance the immune responses towards tumors with poor immunogenicity for cell-based vaccines in mice [39]. In addition, diamine oxidase (PDB ID: 3HIG), which has an activity relationship with sex hormones [40], was ranked as the top 4 binding target. We believe the binding between irinotecan and diamine oxidase (Figure 3b) could impact sex hormones and, consequently, affect libido.

From the above three case studies, we believe our method can not only predict ADRs for drug molecules, but also provide possible mechanism explanations via the binding targets. Since ADRs are complicated and differ from individual to individual [7], such explanation could provide clues for toxicology researchers to generate hypotheses and design wet-lab experiments to understand ADR mechanisms, thus improving the safety evaluation of drugs. Additionally, since our method only requires the structural information of the drug molecules to predict ADRs, it is feasible to use it in the early drug development stage when other types of information about drug candidates are limited.

3.2.4. Feature-based analysis for cataract subcapsular

Cataract subcapsular had the top 6 AUROC value out of 1,533 ADR prediction models (Table 1). We analyzed the docking scores from each of the 600 protein features with the label vector of cataract subcapsular in order to evaluate their individual performance. If a protein feature is associated with an ADR, its docking scores corresponding to positive labels are expected to be significantly lower (more energetically favorable) than the docking scores corresponding to the negative labels, as indicated by both the mean docking scores and the computed ANOVA p value. The top 5 associated protein features which fulfill the mean energetic requirement were ranked by ANOVA p values in Table 4. We found that pyridoxal kinase (PDB ID: 4EOH) ranked second among the protein features. It was reported that pyridoxal-aminoguanidine is more effective than aminoguanidine in the prevention of cataract in diabetic rats [41]. Since the function of pyridoxal kinase is to alter pyridoxal levels, it is possible that a drug that affects the activity of pyridoxal kinase can change the pyridoxal level in...
the human body, thus altering patient vulnerability to cataract subcapsular.

From this feature-based analysis, we showed that it is possible to use our method to find protein targets that are associated with ADRs, thus generating hypothesis that may help us to explore and understand the mechanisms of ADRs.

3.3. Discussion of different types of ADR prediction models

There are several different types of machine learning models that can be leveraged to predict ADRs. These include (1) developing separate models for each ADR, as described in this study, and (2) developing only one model to predict all ADRs. For the second approach, researchers need to harvest features for ADRs such that each row in the training set represents a drug-ADR pair and the training set contains both the drug and ADR features. The label for each row is either positive (representing a known drug-ADR association) or negative (representing an unknown drug-ADR association). Studies have utilized this type of model for the prediction of drug repositioning [42] and drug-drug interactions [43]. Different from our approach, this type of method usually utilizes randomly generated associations as negative samples, making it easier to balance the positive/negative class ratio. However, in order to develop this type of model, it is challenging to identify a good feature set for ADRs. We would like to explore this type of prediction models as our future work.

3.4. Comparison with related studies

We compared our work with five related studies [26], [27], [44]--[46] in Table 5. We found that four out of the five studies used features that require more information than just the molecular structures, such as drug targets, indications, gene expressions and other biological and phenotypic properties [26], [44]--[46]. Although these additional features provide useful information about the molecules and improve prediction performance, they are often not available during early drug development stage. LaBute et al. [27] used a similar molecular docking approach which only required structural input; however, they studied a smaller number of ADRs (10 groups of 85 ADRs) and did not focus on providing biological explanations from their prediction. We believe our work has the advantage of predicting ADRs and providing potential biological mechanisms with a reasonable performance. Additionally, it only requires the structural inputs of the molecules. Thus, our method is applicable to safety evaluation for drug molecules both in the development stage and on the market.

CONCLUSIONS

We prepared structures of 1,231 small molecule drugs and 600 unique human proteins and generated binding scores between them using molecular docking. We developed machine learning models using the molecular docking features to predict 1,533 ADRs. The models achieved reasonably good performance with an overall AUROC value of 0.843, outperforming structural fingerprint-based methods. We found that our machine learning models can successfully predict skin striae for fluticasone propionate, mometasone-induced dermatitis acneiform and irinotecan-induced decreased libido, as three case studies. We further analyzed the binding characteristics of proteins that are top ranked or closely associated with the ADRs in order to find possible interpretations of the ADR mechanisms. We believe the machine learning models based on molecular docking features can not only help with ADR prediction for drug molecules, but also have the advantage of providing possible explanations for the underlying mechanisms of ADRs.

LIST OF ABBREVIATIONS

ADR: adverse drug reaction; ANOVA: Analysis of Variance; AUPR: area under the precision-recall curve; AUROC: area under the receiver operating characteristic curve; CPI: chemical-protein interactome; CD40L: cytokine Cluster of Differentiation 40 ligand; FAERS: FDA Adverse Event Reporting System (FAERS); ECFP: extended connectivity fingerprints; FCFP: functional-class fingerprints; FDA: US Food and Drug Administration; FP4: Fingerprint 4; MACCS: Molecular ACCess System; ROC: receiver operating characteristic curve; RORγt: orphan nuclear receptor gamma; PDB: Protein Data Bank; TNF: tumor necrosis factor.

CONFLICT OF INTEREST

HL, AFN, JH, and PZ are employees of IBM T.J. Watson Research Center. LY is a current employee at Bayer Pharma AG. However, this study was based on his previous work at Shanghai Jiao Tong University and Bayer Pharma AG was not involved.

ACKNOWLEDGEMENTS

The authors would like to thank Zhaonan Sun, Kenney Ng, Harry Stavropoulos, Wendy Cornell and Eileen Koski for their support during the progression of this study.

AUTHOR CONTRIBUTIONS

HL and PZ designed and led the project. HL and PZ collected the data and implemented the methods. HL, AF, NS, LY, JH, and PZ discussed the methods, analyzed the results, and prepared case studies. HL and PZ wrote the manuscript. All authors read and approved the final manuscript.

SUPPLEMENTARY MATERIALS

Supplementary Table S1. List of 655 drug molecules from DrugBank; Supplementary Table S2. List of 600 protein targets from PDBBind; Supplementary Table S3. List of 1,533 ADR labels from SIDER.

REFERENCES


Molecular Docking for ADR predictions

2016, pp. 523

Intelligent Applications Med.

driven prediction of drug effects and interactions," ARWAR: A network approach for predicting Adverse Drug Application of network theory in predicting potential adverse drug

W46 interactions through implementing the chemical

reactions implementing the

issue, pp. W492

protein i

H. Luo

hypersensitivity," Science,

M. Chen, M. Zhang, J. Borlak,

pp. 201

M. Chen

and J. McCluskey, "Drug Hypersensitivity and Human Leukocyte

clindamycin

Y. Yang

Pharmacogenomics

F. Yang

syndrome/toxic epidermal necrolysis caused by methazolamide in


D.-S. Cao et al., “Integrating Multiple Evidence Sources to Predict Adverse Drug Reactions Based on a Systems Pharmacology


Received: March 20, 201X
Revised: April 16, 201X
Accepted: April 20, 201X
Figure 1. Flow chart of the model training, evaluation and feature analysis process. We collected 1,231 drug molecules from DrugBank (a) and 600 unique human proteins from PDBBind (b) to construct an in-silico chemical-protein interactome (CPI) using molecular docking (c). We collected 1,533 ADRs from SIDER database that have at least 5 or more positive drugs (d). Based on the existing drug-ADR knowledge, machine learning models (e) were trained to predict ADRs based on CPI features. We evaluated the performance of the ADR predictions based on a 10-fold cross-validation framework (f) and analyzed top binding or associated proteins (g) in the experiments.
Figure 2. Overall performance of the prediction models demonstrated by the Receiver Operating Characteristic (ROC) curve (left) and Precision-Recall curve (right) using molecular docking and seven structure fingerprints during 10-fold cross-validations.
Figure 3. The predicted binding conformations (a) between mometasone and the orphan nuclear receptor gamma (ROR\(_\gamma\)) ligand-binding domain (PDB ID: 3B0W) and (b) between irinotecan and diamine oxidase (PDB ID: 3HIG) visualized via JSmol. The backbones of small molecules are shown in black color and the proteins are shown as grey and white rockets. The interacting residues between the small molecules and the proteins are highlighted.
### TABLES

**Table 1** Top performed ADR prediction models ranked by AUROC values

<table>
<thead>
<tr>
<th>CUI</th>
<th>ADR</th>
<th>Accuracy</th>
<th>AUROC</th>
<th>AUPR</th>
<th>F1</th>
<th>Precision</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0948226</td>
<td>Ureteral spasm</td>
<td>0.998</td>
<td>0.998</td>
<td>0.663</td>
<td>0.923</td>
<td>0.857</td>
<td>1.000</td>
<td>0.998</td>
</tr>
<tr>
<td>C0235409</td>
<td>Increased insulin requirement</td>
<td>0.994</td>
<td>0.996</td>
<td>0.557</td>
<td>0.800</td>
<td>0.667</td>
<td>1.000</td>
<td>0.994</td>
</tr>
<tr>
<td>C0085570</td>
<td>Alkalosis hypokalaemic</td>
<td>0.992</td>
<td>0.995</td>
<td>0.502</td>
<td>0.737</td>
<td>0.583</td>
<td>1.000</td>
<td>0.992</td>
</tr>
<tr>
<td>C0013928</td>
<td>Fat embolism</td>
<td>0.989</td>
<td>0.994</td>
<td>0.419</td>
<td>0.667</td>
<td>0.500</td>
<td>1.000</td>
<td>0.989</td>
</tr>
<tr>
<td>C0018813</td>
<td>Myocardial rupture</td>
<td>0.989</td>
<td>0.994</td>
<td>0.419</td>
<td>0.667</td>
<td>0.500</td>
<td>1.000</td>
<td>0.989</td>
</tr>
<tr>
<td>C0235259</td>
<td>Cataract subcapsular</td>
<td>0.985</td>
<td>0.990</td>
<td>0.656</td>
<td>0.762</td>
<td>0.640</td>
<td>0.941</td>
<td>0.986</td>
</tr>
<tr>
<td>C0151865</td>
<td>Pregnancy test false positive</td>
<td>0.978</td>
<td>0.984</td>
<td>0.382</td>
<td>0.400</td>
<td>0.278</td>
<td>0.714</td>
<td>0.980</td>
</tr>
<tr>
<td>C0162311</td>
<td>Androgenetic alopecia</td>
<td>0.970</td>
<td>0.983</td>
<td>0.172</td>
<td>0.333</td>
<td>0.200</td>
<td>1.000</td>
<td>0.970</td>
</tr>
<tr>
<td>C0152010</td>
<td>Withdrawal bleed</td>
<td>0.986</td>
<td>0.975</td>
<td>0.149</td>
<td>0.308</td>
<td>0.250</td>
<td>0.400</td>
<td>0.991</td>
</tr>
<tr>
<td>C1735601</td>
<td>Floppy iris syndrome</td>
<td>0.991</td>
<td>0.965</td>
<td>0.433</td>
<td>0.571</td>
<td>0.667</td>
<td>0.500</td>
<td>0.997</td>
</tr>
</tbody>
</table>
### Table 2 Top predicted binding proteins for mometasone

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Protein name</th>
<th>Binding score</th>
</tr>
</thead>
<tbody>
<tr>
<td>2JB6</td>
<td>Fab fragment Mor03268</td>
<td>-10.9</td>
</tr>
<tr>
<td>2WU6</td>
<td>CDC2-like kinase isoforms 3 (CLK3)</td>
<td>-10.6</td>
</tr>
<tr>
<td>3B0W</td>
<td>Orphan nuclear receptor gamma (RORγt) ligand-binding domain</td>
<td>-10.4</td>
</tr>
</tbody>
</table>
### Table 3 Top predicted binding proteins for irinotecan

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Protein name</th>
<th>Binding score</th>
</tr>
</thead>
<tbody>
<tr>
<td>3LKJ</td>
<td>TNF family cytokine CD40 ligand (CD40L)</td>
<td>-13.4</td>
</tr>
<tr>
<td>4GE7</td>
<td>Kynurenine aminotransferase II</td>
<td>-13.3</td>
</tr>
<tr>
<td>3FUN</td>
<td>Leukotriene A4 hydrolase</td>
<td>-13.2</td>
</tr>
<tr>
<td>3HIG</td>
<td>Human diamine oxidase</td>
<td>-12.6</td>
</tr>
</tbody>
</table>
Table 4 Top 5 associated proteins towards cataract subcapsular ranked by ANOVA p values

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Protein name</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4IJQ</td>
<td>Hypoxanthine phosphoribosyltransferase 1 (HPRT1)</td>
<td>4.29E-11</td>
</tr>
<tr>
<td>4EOH</td>
<td>Pyridoxal (pyridoxine, vitamin B6) kinase (PDXK)</td>
<td>9.23E-11</td>
</tr>
<tr>
<td>2WU6</td>
<td>CDC like kinase 3 (CLK3)</td>
<td>4.63E-10</td>
</tr>
<tr>
<td>4N7U</td>
<td>Butyrophilin subfamily 3 member A1 (BTN3A1)</td>
<td>7.61E-10</td>
</tr>
<tr>
<td>3B0W</td>
<td>RAR related orphan receptor C (RORC)</td>
<td>1.64E-09</td>
</tr>
</tbody>
</table>
Table 5 A comparison between our work and other related studies on ADR prediction. The numbers of molecules, features, numbers of ADRs, label sources and performance values were obtained from the original studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Number of molecules</th>
<th>Features</th>
<th>Only require structures?</th>
<th>Number of ADRs</th>
<th>Label source</th>
<th>AUROC</th>
<th>AUPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our work</td>
<td>655</td>
<td>Molecular docking (600 proteins)</td>
<td>Yes</td>
<td>1,533</td>
<td>SIDER</td>
<td>0.843</td>
<td>0.395</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E-State fingerprints</td>
<td>Yes</td>
<td></td>
<td></td>
<td>0.797</td>
<td>0.362</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ECFP6 fingerprints</td>
<td>Yes</td>
<td></td>
<td></td>
<td>0.794</td>
<td>0.351</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FCFP6 fingerprints</td>
<td>Yes</td>
<td></td>
<td></td>
<td>0.798</td>
<td>0.359</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FP4 fingerprints</td>
<td>Yes</td>
<td></td>
<td></td>
<td>0.806</td>
<td>0.370</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR fingerprints</td>
<td>Yes</td>
<td></td>
<td></td>
<td>0.786</td>
<td>0.341</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MACCS fingerprints</td>
<td>Yes</td>
<td></td>
<td></td>
<td>0.802</td>
<td>0.356</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PubChem fingerprints</td>
<td>Yes</td>
<td></td>
<td></td>
<td>0.810</td>
<td>0.358</td>
</tr>
<tr>
<td>LaBute et al.</td>
<td>560</td>
<td>Molecular docking (409 proteins)</td>
<td>Yes</td>
<td>85 (10 groups)</td>
<td>SIDER</td>
<td>0.60-</td>
<td>0.69</td>
</tr>
<tr>
<td>[27]</td>
<td></td>
<td>Drug targets, indications and ADRs</td>
<td>No</td>
<td>10 (for validation)</td>
<td>SIDER/Electronic Health Records</td>
<td>0.92</td>
<td>-</td>
</tr>
<tr>
<td>Bean et al.</td>
<td>524</td>
<td>Structures, gene expressions and multiple evidences</td>
<td>No</td>
<td>817</td>
<td>SIDER</td>
<td>0.57-</td>
<td>0.88</td>
</tr>
<tr>
<td>[44]</td>
<td></td>
<td>Chemical, biological and phenotypic properties</td>
<td>No</td>
<td>22</td>
<td>SIDER</td>
<td>0.48-</td>
<td>1.00</td>
</tr>
<tr>
<td>Cao et al.</td>
<td>746</td>
<td>Structures, gene expressions and multiple evidences</td>
<td>No</td>
<td>817</td>
<td>SIDER</td>
<td>0.57-</td>
<td>0.88</td>
</tr>
<tr>
<td>Jamal et al.</td>
<td>928</td>
<td>Chemical, biological and phenotypic properties</td>
<td>No</td>
<td>22</td>
<td>SIDER</td>
<td>0.48-</td>
<td>1.00</td>
</tr>
<tr>
<td>Liu et al. [26]</td>
<td>832</td>
<td>Chemical, biological and phenotypic properties</td>
<td>No</td>
<td>1,385</td>
<td>SIDER</td>
<td>0.952</td>
<td>-</td>
</tr>
</tbody>
</table>