Citation: Clin Transl Sci (2016) 9, 311–320; doi:10.1111/cts.12422


ARTICLE

# Estimation of Maximum Recommended Therapeutic Dose Using Predicted Promiscuity and Potency

T Liu[1], T Oprea[2], O Ursu[2], C Hasselgren[3] and RB Altman[4,*]

**We report a simple model that predicts the maximum recommended therapeutic dose (MRTD) of small molecule drugs based on an assessment of likely protein–drug interactions. Previously, we reported methods for computational estimation of drug promiscuity and potency. We used these concepts to build a linear model derived from 238 small molecular drugs to predict MRTD. We applied this model successfully to predict MRTDs for 16 nonsteroidal antiinflammatory drugs (NSAIDs) and 14 antiretroviral drugs. Of note, based on the estimated promiscuity of low-dose drugs (and active chemicals), we identified 83 proteins as "high-risk off-targets" (HROTs) that are often associated with low doses; the evaluation of interactions with HROTs may be useful during early phases of drug discovery. Our model helps explain the MRTD for drugs with severe adverse reactions caused by interactions with HROTs.**
*Clin Transl Sci* (2016) **9**, 311–320; doi:10.1111/cts.12422; published online on 13 October 2016.

### Study Highlights

**WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?**
✔The maximum recommended therapeutic dose (MRTD) estimates the upper limit beyond which a drug's efficacy is not increased and side effects begin to outweigh beneficial effects. Currently, MRTD is empirically derived from human clinical trials. We have an opportunity to use computational methods to study the molecular basis of the MRTD.

**WHAT QUESTION DID THIS STUDY ADDRESS?**
✔What are the factors that affect MRTD estimation? What molecular targets may cause the most undesirable off-target activities?

**WHAT THIS STUDY ADDS TO OUR KNOWLEDGE**
✔We built a simple model that predicts MRTD of small-molecule drugs based on an assessment of likely protein–drug interactions. We found two important factors of MRTD: drug promiscuity and pseudo-potency. We identified 83 proteins as "high-risk off-targets" (HROTs) that may cause most undesirable adverse reactions.

**HOW THIS MIGHT CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE**
✔Our MRTD model reveals some molecular aspects of drug action. The ability to predict MRTD directly from drug target interactions is both scientifically and clinically important in terms of drug development and use. The predicted MRTD can be used to estimate the maximum recommended starting dose (MRSD) when designing phase I human clinical trials. The identification of HROTs provides a novel and reliable set of "red flags" for pharmacological profiling.

The "maximum recommended therapeutic dose" (MRTD) for a drug is the upper limit beyond which efficacy is not increased and side effects begin to outweigh beneficial effects.[1] MRTD is empirically derived from human clinical trials, and provides a threshold for dose-related side effects. The US Food and Drug Administration (FDA) created expert systems that use quantitative structure activity relationship (QSAR) methods to estimate both the MRTD as well as the "no effect level" (NOEL) of organic chemicals in humans.[1,2] These models use data obtained from pharmaceutical clinical trials and postmarket surveillance of the adverse reactions reported in the FDA's Adverse Event Reporting System (AERS) databases.[3] Ideally, MRTD estimates provide a relevant, accurate, sensitive, and specific estimate of the toxic dose level of chemicals in humans.

Estimating the best "first in human" (FIH) dose is also an essential activity in clinical drug development. FIH dose estimation is usually based on the "no observable adverse effect levels" (NOAELs) in multiple species.[4] Agoram reported a relationship between pharmacokinetic profiles and the FIH dose.[5] Other pharmacokinetic models predict human clearance (CL) and bioavailability, with an emphasis on toxicity.[5,6] However, drug effectiveness is not typically the focus when estimating FIH dose.

The predicted MRTD can be used to estimate the maximum recommended starting dose (MRSD) and FIH dose for

[1]Department of Genetics, Stanford University, Stanford, California, USA; [2]Department of Internal Medicine, Translational Informatics Division, University of New Mexico, Albuquerque, New Mexico, USA; [3]PureInfo Discovery Corp., Albuquerque, New Mexico, USA; [4]Departments of Bioengineering and Genetics, Stanford University, Stanford, California, USA. *Correspondence: RB Altman (russ.altman@stanford.edu)
Received 17 May 2016; accepted 1 September 2016; published online on 13 October 2016. doi:10.1111/cts.12422

phase I human clinical trials.[2] However, there are no reported methods for estimating MRTDs in the absence of clinical data. Branham *et al.* associated the chemical properties of antiretroviral drug molecules to their MRTDs.[7] Of the six properties examined, only aqueous solubility and biodegradation probability were statistically associated with MRTDs. The model was limited to 31 antiretroviral drugs and is not directly applicable to other drug classes. Thus, the ability to predict MRTD directly from drug target interactions is both clinically and scientifically attractive for drug development and treatment management.

We have previously reported two methods for computational profiling of drug promiscuity and target druggability.[8,9] Promiscuity is often considered a major factor in determining drug side effects.[10] A drug's promiscuity can be measured by its binding spectrum to (ideally) all human proteins in the cell. A protein's druggability is its ability to be modulated by high-affinity interactions with small-molecule drugs. Although we do not have direct means to predict potency, druggability predictions for the known targets of a drug can be used as a proxy estimate of its potency. In particular, we compute the average druggability of all the known targets of a small molecule—high average druggability implies high average affinity and thus high potency.[8] Conversely, low average druggability implies low average affinity and low potency. For this discussion, we will refer to the average druggability as pseudo-potency to stress that it is not a direct measure of potency (see Methods for detailed definition).

MRTD is an empirical parameter that draws a line between therapeutic (desired) effects and adverse or toxic (undesired) side effects.[11] Drug promiscuity and pseudo-potency contribute to undesired and desired effects, respectively. Therefore, MRTD should be a function of both promiscuity and pseudo-potency. For example, the MRTD of celecoxib is relatively low compared with that of other nonsteroidal antiinflammatory drugs (NSAIDs). Using our computational methods,[8,9] we inspected the binding site of celecoxib in its functional target COX2 (cyclooxygenase 2) and found optimized interactions and high druggability score consistent with high pseudo-potency. At the same time, we also found that celecoxib is a drug of high promiscuity as measured by the number of human proteins to which it may bind. The balance between these two characteristics qualitatively explains celecoxib's relatively low dose: celecoxib achieves its desired effects because of the optimal interactions (and high pseudo-potency) with COX2. On the other hand, its dose cannot be increased very much because high promiscuity leads to side effects.

Although MRTD is established during clinical trials, it can be changed once data from patient exposures are analyzed. Ultimately, this information is reflected in drug labels and guides prescribing physicians. In this work, we aim to develop a simple model of MRTD based on these two molecular attributes, pseudo-potency and promiscuity—both of which can be estimated using basic molecular structure data that is often available to drug developers. Based on this model, we predict and reevaluate the MRTD of drugs with severe side effects and provide insights about target interactions that might best be avoided during drug development.

## METHODS

### Datasets

The *Drug Dataset* comprises 238 small-molecule drugs that satisfy the following standards: (i) The high-quality 3D structures of a drug's binding sites are available in Protein Data Bank (PDB); (ii) The MRTD values are available from the FDA MRTD database (http://www.fda.gov/cder/). We normalized the original MRTD values by two steps: (i) The original MRTD values were normalized to 60 kg body weight; (ii) we renormalized the MRTD values to 70 kg, which is the "average" adult mass is 70 kg in physiology studies.[2,12] We further divided the dosage (expressed in mg) by the molecular weight (MW) of the actual drug—observing the actual drug formulation unless the active substance MW was explicitly stated in the package insert. When multiple values were available, we used the MRTD for the oral formulation. In this study, log(MRTD) refers to the logarithm of MRTD, expressed in $\mu$Mol/kg/day. The values of MRTD of the 234 drugs range from 10e-5 to 10e4 $\mu$Mol/kg/day, representing the complete MRTD range of the FDA database (**Supplementary Figure S1**).

The *Human Protein Dataset* comprises 2,291 proteins from a nonredundant representative set (90% identity) of human proteins. We used the following filters: (i) a high-quality 3D structure (x-ray resolution higher than 2.5 Å) is available in PDB. (ii) The structure is cocrystalized with a small molecule ligand.[13] Using these criteria, we collected 46 low-dose drugs (MRTD $<1$ $\mu$Mol/kg/day) and 37 high-dose drugs (MRTD $>100$ $\mu$Mol/kg/day) from the 238 drugs (**Table S1** and **Figure S1**).

### Predict promiscuity

Given a drug, we predict its probability of binding to all proteins in the *Human Protein Dataset* (Section 1). We employ a previously reported method, PocketFEATURE,[8] which computes the structural similarity between two binding sites in order to calculate the probability that a drug binding one site also binds the other. More similar sites are more likely to share drug binding profiles. We describe a drug by enumerating its binding microenvironments (physiochemical and structural properties) in a target protein using the FEATURE system.[14] FEATURE calculates a set of 80 physicochemical properties collected over six concentric spherical shells (total 480 properties $=$ 80 properties $\times$ 6 shells) centered on the predefined functional center. PocketFEATURE uses the FEATURE representation to calculate site similarities by aligning microenvironments between two sites. A more negative score suggests binding site similarity and thus a higher probability of drug binding to a site similar to its known binding site. A cutoff of –2.0 indicates likely binding of a drug to a protein target. A more stringent cutoff –2.5 indicates likely more specific binding. The similarity between a drug's binding site and each of the 2,291 binding sites in the *Human Protein Dataset* can be calculated by PocketFEATURE. Given a drug, we then count the number of proteins in the *Human Protein Dataset* that are predicted to bind the drug (using cutoff of –2.0). We then calculate the average pseudo-affinity as an indication of promiscuity of the drug.

### Estimate pseudo-potency

We previously reported a method, DrugFEATURE, that evaluates a protein's potential to bind drug-like molecules by assessing the microenvironments in putative binding sites.[8] DrugFEATURE estimates the potential for high binding affinity between a drug and a protein. Given a drug, the average druggability of its functional targets bound is a proxy measure of the drug's binding affinity in these targets, and measures the amount of drug required to modulate the target. In this work, we call the average druggability the "pseudo-potency." We apply DrugFEATURE to compute pseudo-potency for 234 drugs. For each drug, we collect its functional targets from DrugBank and seek cocrystallized structures of the targets. We then estimate the drug's pseudo-potency by averaging the druggability of each of the target binding sites.

### Build linear models and predict MRTD

We built a linear model for MRTD, using independent variables promiscuity and pseudo-potency using the R package (Vienna, Austria). We analyzed the significance of each variable by analysis of variance (ANOVA). We employed leave-one-out crossvalidation to build linear models and predict MRTDs of 14 antiretroviral drugs and 16 NSAIDs.

### Identify high-risk off-targets (HROTs)

From the 2,291 proteins, we identified 83 targets that (i) do not bind any of the 37 high-dose drugs (PocketFEATURE stringent cutoff –2.5) and (ii) bind to at least 5 of the 46 high-dose drugs. We evaluate the statistical significance of targets matching these criteria using the hypergeometric distribution over the 238 drugs. The probability of observing binding to low-dose drugs is calculated as:

$$pl = f(xl|M, K, Nl) = \frac{\binom{K}{xl}\binom{M-K}{Nl-xl}}{\binom{M}{Nl}}$$

Where M is the size of the population (238 drugs); K is the number of drugs that bind to the given target; Nl is the size of samples drawn (46 low-dose drugs) and xl is the number of bindings observed in low-dose drugs. The probability of not binding to any of high-dose drugs is calculated as:

$$ph = f(xh|M, K, Nh) = \frac{\binom{K}{xh}\binom{M-K}{Nh-xh}}{\binom{M}{Nh}}$$

Where M is the size of the population (238 drugs), K is the number of drugs that bind to the given target, Nh is the size of samples drawn (37 high-dose drugs), and xh is zero (no observed binding). The significance of a given target is calculated as $P = pl \times ph$.

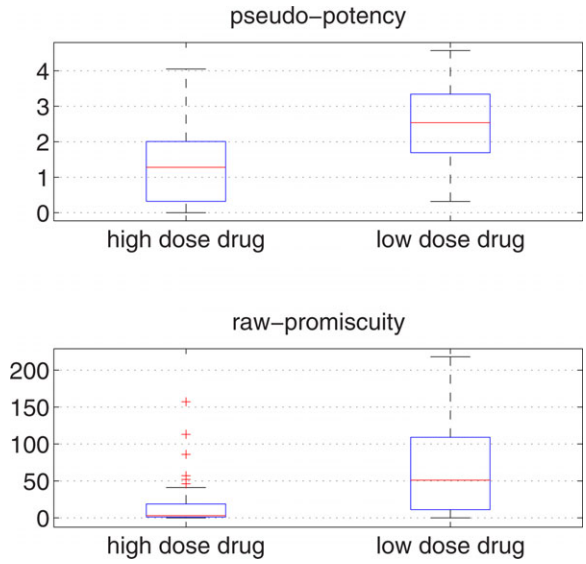


**Figure 1** We compare the pseudo-potency and promiscuity of 37 high-dose (MRTD >100 $\mu$Mol/kg/day) and 46 low-dose drugs (MRTD <1 $\mu$Mol/kg/day). The boxplot shows that high-dose and low-dose drugs have significantly different pseudo-potency (*t*-test *P*-value = 2.24e-4) and raw promiscuity (*t*-test *P*-value = 3.29e-5). We displayed the distribution of data noting the minimum, first quartile, median, third quartile, and maximum.

## RESULTS

### Pseudo-potency and promiscuity are two factors of MRTD

We employed PocketFEATURE[8] to predict affinity between each of the 238 drugs in the *Drug Dataset* and the 2,291 proteins in the *Human Protein Dataset* (see Methods). We have previously shown that the accuracy of PocketFEATURE is reasonably good.[10,15] The predicted scores approximate the probability of binding between a drug and a protein, and therefore the set of the predicted affinity scores between a drug and the 2,291 human proteins (*Human Protein Dataset)* can be used as an estimate of drug promiscuity. We also estimated the pseudo-potency of the 238 drugs in *Drug Dataset* by averaging the druggability of their functional targets, using the DrugFEATURE algorithm.[9]

For the 238 drugs in this study, there were 37 high-dose drugs (MRTD >100 $\mu$Mol/kg/day) and 46 low-dose drugs (MRTD <1 $\mu$Mol/kg/day) (**Figure S1**). We compared the pseudo-potency and the raw scores of promiscuity of the 37 high-dose and the 46 low-dose drugs (**Figure 1**). The average pseudo-potency of high-dose drugs is 1.2 and that of low-dose drugs is 2.6. The average promiscuity of high-dose drugs is lower than that of low-dose drugs. High- and low-dose drugs have significantly different pseudo-potency (*P*-value = 2.24e-4) and raw promiscuity (*P*-value = 3.29e-5).

### Linear model

We built a predictive model for MRTD based on promiscuity and pseudo-potency. The results of the multiple linear

**Table 1** Multiple linear regression analysis. The linear regression model is: Log(MRTD) ~ pseudo-potency + promiscuity. Panel A shows that the model is significant, with F-statistic of the linear fit vs. the constant model is 8.098 (*P*-value of 3.97e-4). The R-squared value (0.065) indicates that the model explains 6.5% of the variability in the response. Panel B shows that F-statistics for assessing the statistical significance of pseudo-potency and promiscuity. Both promiscuity and pseudo-potency contribute to the model of MRTD. The ANOVA table shows that combining pseudo-potency and promiscuity improves the model

**A. Estimated coefficients**

| | Estimate | Std error | *t* value | Pr(>\|t\|) |
|---|---|---|---|---|
| Intercept | 4.2702 | 0.6254 | 6.828 | 7.28e-11*** |
| Promiscuity | 0.4640 | 0.1978 | 2.345 | 0.01985* |
| Pseudo potency | −0.3217 | 0.1134 | −2.837 | 0.00496** |

**B. ANOVA table of the model terms**

| | Sum sq | Mean sq | F-value | Pr(>F) |
|---|---|---|---|---|
| Promiscuity | 55.37 | 55.373 | 8.1488 | 0.004694** |
| Pseudo-potency | 54.68 | 54.682 | 8.0471 | 0.004956** |

regression analysis are in **Table 1**. Our model for MRTD is:

$$\log(MRTD) = 0.4640 \times Promiscuity - 0.3217 \times PseudoPotency + 4.2702$$

Panel A of **Table 1** shows that the model is significant; the F-statistic of the linear fit vs. the constant model is 8.098 (*P*-value of 3.97e-4). The R-squared value (0.065) shows that the model explains only 6.5% of the variability in the response. Panel B shows the F-statistic for assessing the statistical significance of pseudo-potency and promiscuity in the model; both variables significantly contribute to the prediction. The ANOVA table also shows that combining pseudo-potency with promiscuity improves the model.

### Predicted MRTD

We evaluated prediction power using leave-one-out cross-validation. We also analyzed prediction results on two important pharmaceutical categories: 14 antiretroviral drugs and 16 NSAIDs. **Figure 2** shows that the predicted values correlate with the known MRTDs for both categories. The correlation between the predicted values and the known MRTD for the 16 NSAIDs is 0.5 (*P*-value 0.07). The correlation between the predicted values and the known MRTD for the 14 antiretroviral drugs is 0.9 (*P*-value 0.0001). Drugs that achieved good prediction performance (within one unit of log(MRTD)) are highlighted in red, including eight antiretroviral drugs and seven NSAIDs. All seven NSAIDs with good performance are slow time-dependent inhibitors. Six of the nine NSAIDs for which predictions did not achieve good performance are rapid inhibitors (**Supplementary Table 2A**). The eight antiretroviral drugs with accurate predictions consist of two nucleoside reverse transcriptase inhibitors (NRTI), one non-nucleoside reverse transcriptase inhibitor (NNRTI), and five HIV protease inhibitors (HIV-PI) (**Supplementary Table 2B**). Drugs of high dose, such as foscarnet and zidovudine, tend to have less accurate predictions.

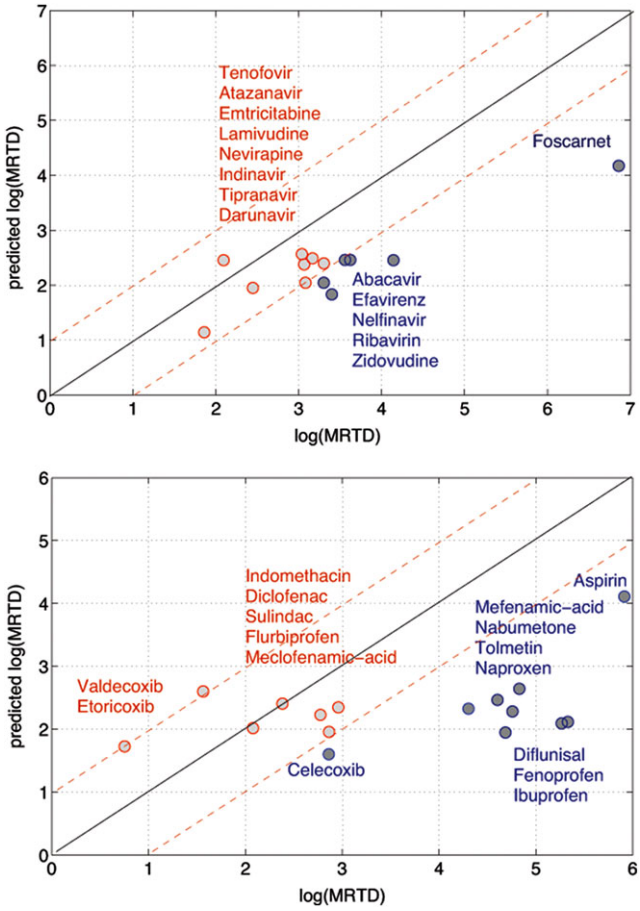


**Figure 2** Predict MRTD. The top panel shows 16 NSAIDs and the bottom panel shows 14 antiretroviral drugs. MRTD is in $\mu$Mol/kg/day. The x-axis is the known MRTD and the y-axis is the prediction. The dotted lines suggest prediction errors of one unit of log(MRTD). Predictions within the dotted lines are considered good performance and highlighted in red. The correlation between the predicted values and the known MRTD for the 16 NSAIDs is 0.5 (*P*-value 0.07). The correlation between the predicted values and the known MRTD for the 14 antiretroviral drugs is 0.9 (*P*-value 0.0001).

### High-risk off-targets (HROTs)

We identified 83 proteins that are predicted to bind low-dose drugs more frequently than high-dose drugs. We call these proteins "high-risk off-targets" because they are associated with drugs whose dose is limited by high promiscuity and low potency. They consist of 32 proteins that are associated with transcription process, 36 receptor-related proteins, and 20 hormone receptors. There are six G-protein-coupled receptors (GPCRs, **Table 2**).

## DISCUSSION

### Pseudo-potency and promiscuity leverage a drug's MRTD

The MRTD represents the margin between the desired effects and adverse reactions. It is not surprising that MRTD is associated with drug promiscuity, which is an important factor for drug adverse reactions. However, the number of drug

**Table 2** High-risk off-targets. We identified 83 HROTs that bind to low-dose drugs, but not high-dose drugs. The first column lists the UniProt IDs, the third lists gene names, and the fourth one lists key GO terms for the proteins. The second column shows the hypergeometric *P*-value computation for the significance of extracting HROTs. A total of 32 proteins are associated with transcription process; 36 are receptors and 20 are hormone receptors. There are six proteins involved in GPCR pathways. In addition, 11 enzymes involved in key pharmacokinetics are marked with an asterisk

| UniProt | *P*-value | Gene name | Key go terms |
|---|---|---|---|
| Q15788 | 4.214e-06 | NCOA1 BHLHE74 SRC1 | androgen receptor binding<br>nuclear hormone receptor binding<br>transcription coactivator activity |
| Q96RI1 | 4.214e-06 | NR1H4 BAR FXR HRR1 RIP14 | bile acid binding<br>transcription factor activity<br>ligand-activated sequence-specific DNA binding |
| Q9Y2Q3 | 5.4471e-06 | GSTK1 HDCMD47P | glutathione peroxidase activity<br>receptor binding |
| P41595 | 8.0883e-06 | HTR2B | G-protein alpha-subunit binding<br>GTPase activator activity<br>serotonin receptor activity |
| Q6VVX0 | 9.0829e-06 | CYP2R1* | heme binding oxidoreductase activity, steroid hydroxylase activity |
| O95749 | 1.145e-05 | GGPS1 | farnesyltranstransferase activity |
| O00482 | 1.7458e-05 | NR5A2 B1F CPF FTF | chromatin binding<br>transcription factor activity<br>ligand-activated sequence-specific DNA binding |
| P35222 | 1.7458e-05 | CTNNB1 CTNNB<br>OK/SW-cl.35 PRO2286 | alpha-catenin binding<br>nuclear hormone receptor binding<br>transcription factor activity |
| Q15466 | 1.98e-05 | NR0B2 SHP | steroid hormone receptor activity<br>transcription factor activity, sequence-specific DNA binding |
| Q13133 | 2.3358e-05 | NR1H3 LXRA | cholesterol binding<br>sterol response element binding<br>transcription coactivator activity |
| P29016 | 3.5549e-05 | CD1B | beta-2-microglobulin binding<br>endogenous lipid antigen binding |
| P29017 | 3.5549e-05 | CD1C | beta-2-microglobulin binding<br>exogenous lipid antigen binding |
| P61769 | 3.5549e-05 | B2M CDABP0092 HDCMA22P | glycoprotein binding<br>identical protein binding |
| Q07869 | 3.8416e-05 | PPARA NR1C1 PPAR | RNA polymerase II transcription factor activity,<br>ligand-activated sequence-specific DNA binding |
| P41146 | 4.129e-05 | OPRL1 OOR ORL1 | G-protein-coupled receptor activity<br>neuropeptide binding |
| P37231 | 4.5007e-05 | PPARG NR1C3 | activating transcription factor binding<br>ligand-dependent nuclear receptor transcription coactivator activity<br>prostaglandin receptor activity<br>retinoid X receptor binding |
| P19793 | 4.6483e-05 | RXRA NR2B1 | 9-cis retinoic acid receptor activity<br>transcription factor activity, sequence-specific DNA binding<br>vitamin D receptor binding |
| P55055 | 4.6483e-05 | NR1H2 LXRB NER UNR | apolipoprotein A-I receptor binding<br>ATPase binding<br>RNA polymerase II transcription factor activity |
| Q15596 | 4.6483e-05 | NCOA2 BHLHE75 SRC2 TIF2 | chromatin binding<br>histone acetyltransferase activity<br>nuclear hormone receptor binding<br>thyroid hormone receptor coactivator activity<br>transcription coactivator activity |
| P08684 | 5.1929e-05 | CYP3A4 CYP3A3* | oxidoreductase activity<br>steroid hydroxylase activity |
| P41145 | 5.8559e-05 | OPRK1 OPRK | dynorphin receptor activity<br>neuropeptide binding<br>opioid receptor activity |
| P51449 | 7.934e-05 | RORC NR1F3 RORG RZRG | steroid hormone receptor activity<br>transcription factor activity, sequence-specific DNA binding |

(*Continued*)

**Table 2** Continued

| UniProt | *P*-value | Gene name | Key go terms |
|---|---|---|---|
| P11597 | 8.1254e-05 | CETP | cholesterol binding<br>cholesterol transporter activity<br>phospholipid transporter activity |
| P10635 | 9.9326e-05 | CYP2D6 CYP2DL1* | arachidonic acid epoxygenase activity oxidoreductase activity<br>steroid hydroxylase activity |
| P04150 | 0.00011289 | NR3C1 GRL | glucocorticoid-activated RNA polymerase II transcription factor activity |
| Q07817 | 0.00020485 | BCL2L1 BCL2L BCLX | protein kinase binding |
| P21453 | 0.00021589 | S1PR1 CHEDG1 EDG1 | G-protein-coupled receptor activity |
| P62508 | 0.00021589 | ESRRG ERR3 ERRG2 | retinoic acid receptor activity transcriptional activator activity, |
| P10276 | 0.00022192 | RARA NR1B1 | retinoic acid receptor activity<br>transcription factor activity |
| P29274 | 0.00022192 | ADORA2A ADORA2 | G-protein-coupled adenosine receptor activity |
| P10827 | 0.0002477 | THRA EAR7 ERBA1 NR1A1 THRA1 THRA2 | chromatin DNA binding<br>steroid hormone receptor activity<br>thyroid hormone receptor activity<br>transcription factor activity |
| Q9H227 | 0.00025318 | GBA3 CBG CBGL1 | beta-galactosidase activity<br>glycosylceramidase activity |
| P09960 | 0.00027014 | LTA4H LTA4 | aminopeptidase activity<br>leukotriene-A4 hydrolase activity |
| P22680 | 0.00028977 | CYP7A1 CYP7* | cholesterol 7-alpha-monooxygenase activity |
| P11511 | 0.00031799 | CYP19A1 ARO1* | oxidoreductase activity |
| Q14994 | 0.00045613 | NR1I3 CAR | androgen receptor activity<br>thyroid hormone receptor activity<br>transcription factor activity |
| P14902 | 0.00051981 | IDO1 IDO INDO | electron carrier activity<br>indoleamine 2,3-dioxygenase activity |
| P27986 | 0.00054914 | PIK3R1 GRB1 | 1-phosphatidylinositol-3-kinase regulator activity<br>insulin-like growth factor receptor binding transcription factor binding<br>transmembrane receptor protein tyrosine kinase adaptor activity |
| P11509 | 0.00068414 | CYP2A6 CYP2A3* | arachidonic acid epoxygenase activity oxidoreductase activity,<br>steroid hydroxylase activity |
| P10275 | 0.00077224 | AR DHTR NR3C4 | androgen binding<br>ATPase binding<br>transcription factor activity |
| Q9UBK2 | 0.00077224 | PPARGC1A LEM6 PGC1 PGC1A PPARGC1 | androgen receptor binding<br>ligand-dependent nuclear receptor binding<br>ligand-dependent nuclear receptor transcription coactivator activity |
| P05093 | 0.00081711 | CYP17A1 CYP17* S17AH | 17-alpha-hydroxyprogesterone aldolase activity steroid 17-alpha-monooxygenase activity |
| P10826 | 0.0010141 | RARB HAP NR1B2 | RNA polymerase II regulatory region binding<br>steroid hormone receptor activity |
| P11712 | 0.0010293 | CYP2C9 CYP2C10* | drug binding<br>monooxygenase activity<br>steroid hydroxylase activity |
| Q5SQI0 | 0.0010715 | ATAT1 C6orf134 MEC17 Nbla00487 | coenzyme binding [GO:0050662]; tubulin N-acetyltransferase activity [GO:0019799] |
| P06132 | 0.0011377 | UROD | ferrous iron binding [GO:0008198]; uroporphyrinogen decarboxylase activity [GO:0004853] |
| Q9Y6Q9 | 0.00116 | NCOA3 AIB1 BHLHE42 RAC3 TRAM1 | androgen receptor binding<br>nuclear hormone receptor binding<br>thyroid hormone receptor binding<br>transcription coactivator activity |
| O43617 | 0.0011879 | TRAPPC3 BET3 CDABP0066 | TRAPPC3 BET3 CDABP0066 |

(*Continued*)

**Table 2** Continued

| UniProt | *P*-value | Gene name | Key go terms |
|---|---|---|---|
| Q99835 | 0.0012328 | SMO SMOH | G-protein-coupled receptor activity<br>Wnt-protein binding |
| O75469 | 0.0013221 | NR1I2 PXR | drug binding<br>steroid hormone receptor activity<br>transcriptional activator activity |
| O76074 | 0.0014457 | PDE5A PDE5 | 3',5'-cyclic-GMP phosphodiesterase activity<br>phosphodiesterase activity<br>cGMP binding |
| P51160 | 0.0014457 | PDE6C PDEA2 | 3',5'-cyclic-GMP phosphodiesterase activity [GO:0047555];<br>cGMP binding [GO:0030553]; metal ion binding [GO:0046872] |
| P03372 | 0.0016219 | ESR1 ESR NR3A1 | ATPase binding<br>estrogen receptor activity<br>transcription factor activity<br>steroid hormone receptor activity s |
| P12821 | 0.0017773 | ACE DCP DCP1 | actin binding<br>carboxypeptidase activity<br>endopeptidase activity<br>mitogen-activated protein kinase binding |
| P27815 | 0.001848 | PDE4A DPDE2 | 3',5'-cyclic-AMP phosphodiesterase activitycAMP binding |
| O15217 | 0.001859 | GSTA4 | glutathione transferase activity |
| P20813 | 0.0021009 | CYP2B6* | steroid hydroxylase activity |
| Q15119 | 0.0021009 | PDK2 PDHK2 | ATP binding<br>protein kinase activity |
| Q16678 | 0.0021009 | CYP1B1* | aromatase activity |
| P11474 | 0.0023382 | ESRRA ERR1 ESRL1 NR3B1 | steroid hormone receptor activity<br>transcriptional activator activity |
| Q13772 | 0.0023382 | NCOA4 ARA70 ELE1 RFG | androgen receptor binding<br>transcription coactivator activity |
| P33261 | 0.0023533 | CYP2C19* | arachidonic acid epoxygenase activity<br>steroid hydroxylase activity |
| Q92731 | 0.0025393 | ESR2 ESTRB NR3A2 | estrogen receptor activity<br>steroid hormone receptor activity<br>transcription coactivator activity |
| Q86YN6 | 0.0027512 | PPARGC1B PERC PGC1<br>PGC1B PPARGC1 | estrogen receptor binding<br>ligand-dependent nuclear receptor transcription coactivator<br>activity |
| P63092[B] | 0.0027575 | GNAS GNAS1 | GTPase activity<br>signal transducer activity |
| P07550 | 0.0027575 | ADRB2 ADRB2R B2AR | beta2-adrenergic receptor activity<br>epinephrine binding<br>norepinephrine binding<br>potassium channel regulator activity |
| P27487 | 0.0027575 | DPP4 ADCP2 CD26 | dipeptidyl-peptidase activity<br>protease binding<br>virus receptor activity |
| P59768[B] | 0.0027575 | GNG2 | G-protein beta-subunit binding<br>GTPase activity<br>signal transducer activity |
| Q03181 | 0.0027809 | PPARD NR1C2 PPARB | steroid hormone receptor activity<br>transcription factor activity |
| Q9HCD5 | 0.0027809 | NCOA5 KIAA1637 | chromatin binding<br>poly(A) RNA binding |
| P02751 | 0.0029362 | FN1 FN | collagen binding<br>integrin binding<br>peptidase activator activity |
| P05108 | 0.0038778 | CYP11A1 CYP11A* | cholesterol monooxygenase activity<br>iron ion binding |
| P10109 | 0.0038778 | FDX1 ADX | electron carrier activity |

(*Continued*)

**Table 2** Continued

| UniProt | *P*-value | Gene name | Key go terms |
|---|---|---|---|
| P41235 | 0.0047816 | HNF4A HNF4 NR2A1 TCF14 | steroid hormone receptor activity<br>transcription factor activity |
| P55789 | 0.0055894 | GFER ALR HERV1 HPO | flavin adenine dinucleotide binding<br>protein disulfide oxidoreductase activity |
| P68871 | 0.0055894 | HBB | oxygen transporter activity |
| P69905 | 0.0055894 | HBA1; HBA2 | oxygen transporter activity |
| P69891 | 0.0056218 | HBG1 PRO2979 | oxygen transporter activity |
| P28222 | 0.0060852 | HTR1B HTR1DB | serotonin receptor activity |
| P02753 | 0.0062538 | RBP4 PRO2222 | retinol transporter activity |
| Q14541 | 0.0067333 | HNF4G NR2A2 | steroid hormone receptor activity<br>transcription factor activity |
| P14061 | 0.0068003 | HSD17B1 E17KSR EDH17B1 EDH17B2 EDHB17 SDR28C1 | catalytic activity<br>estradiol 17-beta-dehydrogenase activity testosterone dehydrogenase (NAD+) activity |
| P28702 | 0.0068003 | RXRB NR2B2 | 9-cis retinoic acid receptor activity<br>steroid hormone receptor activity<br>transcription factor activity |

Proteins marked with "B" are 3D structures are proteins from bovin (percentage of sequence identities between human and bovin proteins are 99.75% for P63092, 100% for P63212). Key enzymes involved in pharmacokinetics are marked with an asterisk.

adverse reactions alone does not correlate well with MRTD (Figure S2). Therefore, we added drug potency (pseudo-potency) to our model. Thus, we used computational estimates based on 3D structure interactions that were proxies for promiscuity (using PocketFEATURE[8]) and potency (using DrugFEATURE[9]). We have previously shown that the promiscuity is associated with drug adverse reactions.[10] As expected, we found that our estimated pseudo-potency is associated in MRTD (**Figure 1**). Drugs of low MRTD often have high pseudo-potency, since only drugs of high pseudo-potency can achieve their desired effects at low dose. On the other hand, drugs of low pseudo-potency need high dose to reach the desired therapeutic effects. ANOVA test shows that including pseudo-potency improves the linear model, demonstrating that promiscuity and pseudo-potency provide independent, complementary information.

### Predict and reevaluate MRTD

Although our model has a low R-squared value, it shows statistically significant coefficients. Without question, a low R-squared can be problematic when precise estimates are required. However, we have found it useful in two drug categories: 14 antiretroviral drugs and 16 NSAIDs.

For NSAIDs, binding to functional targets COX2 or COX1 (cyclooxygenase 1 and 2, respectively) is a key step of drug action. There are two types of kinetics in NSAIDs binding (see **Supplementary Materials**). One type is rapid binding, including reversible and irreversible inhibitors (e.g., ibuprofen, piroxicam, mefenamic acid, aspirin). The other is slow, time-dependent binding (e.g., celecoxib, diclofenac, flurbiprofen, indomethacin), which often results in higher *in vitro* potency.[16,17] The model achieves better performance on slow time-dependent inhibitors. For rapid inhibitors, the predicted values are often lower than the known MRTDs. Since our model is a simplified one, we do not include other factors that may affect MRTD, such as pharmacokinetic properties, drug metabolizing, enzyme kinetics, and transporter effects. However, it seems that promiscuity estimation based on high-affinity drug-binding sites tends to achieve a higher accuracy, resulting in better performance of MRTD predictions.

For the group of antiretroviral drugs, the clinically effective dose is often accompanied by substantial adverse effects. Our predicted MRTDs are generally lower than the observed MRTDs because most antiretroviral drugs have high promiscuity, corresponding to their severe side effects. However, drugs with high pseudo-potency may be able to exert their desired functions at lower doses. Thus, we propose that the MRTD for those drugs may merit reevaluation. For example, efavirenz and abacavir have high pseudo-potency (**Supplementary Table S2B**). Our predicted MRTD is lower than the empirical MRTD for these two drugs. They may be able to achieve their effects at a dose lower than MRTD because of their high pseudo-potency. Furthermore, both efavirenz and abacavir interact with HROTs (efavirenz interacts with dipeptidyl peptidases and retinol-binding protein; abacavir interacts with dipeptidyl peptidases and T-cell surface glycoprotein CD1b. **Table 2**). Therefore, lowering doses of these two drugs may help reduce the undesired side effects.

### High-risk off-targets (HROTs)

A low-dose drug often has higher potency and higher promiscuity, compared with high-dose drugs (**Figure 1**). In order to identify proteins that may modulate common and important side effects, we seek proteins that frequently interact with low-dose drugs, but not with high-dose drugs. These proteins and their related pathways are sensitive to the modulations induced by drug binding, which could contribute to low tolerance. We defined HROTs as proteins that seem to be dose limiting based on frequent predicted interactions with low-dose drugs.

For example, as a low-dose drug, dexamethasone (0.3 $\mu$Mol/kg/day) is predicted to be highly effective (pseudo-potency score 3.55) and promiscuous (promiscuity score

–3.43). One of its off-targets is glucocorticoid nuclear receptor 2 (NCOA2), which has been identified as an HROT. The binding between NCOA2 and dexamethasone has also been observed experimentally ($IC_{50}$–22 $\mu$M, data from ChEMBL[18]). Meanwhile, NCOA2 has been associated with severe adverse reactions caused by dexamethasone, including menstrual irregularities, cardiomegaly, and cardiac arrest.[10] Active chemicals that interact with HROTs may cause severe side effects, resulting in low tolerance.

In previous work, we identified 50 essential proteins that are significantly associated with drug adverse reactions.[10] Among the 50 essential proteins, nuclear receptors are enriched, suggesting that hormone modulation can contribute to adverse reactions. In this work, among the 83 HROTs nearly half of them are hormone receptors or nuclear receptors (**Table 2**). Another important group are proteins involved in transcription and signaling process, including 32 proteins associated with transcription process. Elucidating interactions between drugs and transcription factors remains a challenge because of the complexity of cellular responses to drugs. Our predictions provide high-risk alerts that drug binding to these genes may cause severe adverse reactions.

Bowes *et al.* have published a "minimal panel" of targets that should be used for pharmacological profiling to identify the most undesirable off-target activities.[19] (The original source of this panel was four major pharmaceutical companies.) These targets often have a high hit rate and a high impact in *in vitro* profiling. They include 24 GPCRs, seven ion channel targets, six intracellular enzymes, three neurotransmitter transporters, two nuclear hormone receptors, and one kinase. We have found that eight HROTs (five GPCRs and three nuclear receptors) overlap with the "minimal panel," suggesting our list of HROTs provides complementary information for interpreting pharmacological profiling. In addition, our HROTs list includes 11 key enzymes in pharmacokinetics (**Table 2**).

### Reliability of computational profiling

In this work we employed two computational predictions as proxies for drug potency and promiscuity. Ideally, these should be derived from direct experimental assays. For drug promiscuity, it would require a complete binding profile between small molecular drugs and a broad spectrum of human proteins (not limited to known drug targets). However, such large-scale binding assays are difficult and expensive. We have inspected the high confidence datasets from ChEMBL[18] and BindingDB[13] and found that on average there are 15 unique assays for each drug. In addition, these assays are biased towards known target proteins. To estimate drug promiscuity, we are also interested in proteins that have not traditionally been considered drug targets. Computational estimation is imperfect, but it can create an unbiased profile of drug binding to a broad spectrum of proteins. Furthermore, our validations have shown that the predicted affinities and experimental assays are well correlated.[10]

The other term, pseudo-potency, is based on our published methods for estimating druggability, which have been validated[9] by nuclear magnetic resonance (NMR) experimental results and comparison to drug discovery outcomes.[20,21] To our knowledge, we are the first to associate these two computational terms, which provide molecular insights into the drug target space and the influences on therapeutic doses. It is clear that there are other factors that contribute to the MRTD, such as clearance and bioavailability. However, data relevant to these factors are limited because they often rely on difficult and expensive *in vivo* preclinical assays as well as *in vitro* metabolism and disposition measurements. Therefore, we have chosen to model only two factors for small molecule drugs and demonstrated their potential to reveal molecular mechanisms of drug actions.

**Acknowledgments.** This work was supported by NIH GM102365, GM072970, U54 HL117798, and U54 NCI 189205. We thank A. Gottlieb, S. Rensi, and Y. Li for helpful discussion.

**Author Contributions.** R.B.A. and T.L. wrote the article; R.B.A. and T.L. designed the research; T.L. performed the research; T.L. analyzed the data; R.B.A., T.O., O.U., and C.H. contributed new reagents/analytical tools.

**Conflict of Interest.** The authors declare no competing financial interest.

1. Matthews, E.J., Kruhlak, N.L., Benz, R.D. & Contrera, J.F. Assessment of the health effects of chemicals in humans: I. QSAR estimation of the maximum recommended therapeutic dose (MRTD) and no effect level (NOEL) of organic chemicals based on clinical trial data. *Curr. Drug. Discov. Technol.* **1**, 61–76 (2004).
2. Contrera, J.F., Matthews, E.J., Kruhlak, N.L. & Benz, R.D. Estimating the safe starting dose in phase I clinical trials and no observed effect level based on QSAR modeling of the human maximum recommended daily dose. *Regul. Toxicol. Pharmacol.: RTP* **40**, 185–206 (2004).
3. Ahmad, S.R. Adverse drug event monitoring at the Food and Drug Administration. *J. Gen. Intern. Med.* **18**, 57–60 (2003).
4. Zou, P. *et al.* Applications of human pharmacokinetic prediction in first-in-human dose estimation. *AAPS J.* **14**, 262–281 (2012).
5. Agoram, B.M. Use of pharmacokinetic/ pharmacodynamic modelling for starting dose selection in first-in-human trials of high-risk biologics. *Br. J. Clin. Pharmacol.* **67**, 153–60 (2009).
6. Buckley, L.A. & Dorato, M.A. High dose selection in general toxicity studies for drug development: A pharmaceutical industry perspective. *Regul. Toxicol. Pharmacol.: RTP* **54**, 301–7 (2009).
7. Branham, M.L., Ross, E.A. & Govender, T. Predictive models for maximum recommended therapeutic dose of antiretroviral drugs. *Comput. Math. Methods Med.* **2012**, 469769 (2012).
8. Liu, T. & Altman, R.B. Using multiple microenvironments to find similar ligand-binding sites: application to kinase inhibitor binding. *PLoS Comput. Biol.* **7**, e1002326 (2011).
9. Liu, T. & Altman, R.B. Identifying druggable targets by protein microenvironments matching: application to transcription factors. *CPT Pharmacometrics Syst. Pharmacol.* **3**, e93 (2014).
10. Liu, T. & Altman, R.B. Relating essential proteins to drug side-effects using canonical component analysis: a structure-based approach. *J. Chem. Inform. Model.* **55**, 1483–1494 (2015).
11. Matthews, E.J., Kruhlak, N.L., Weaver, J.L., Benz, R.D. & Contrera, J.F. Assessment of the health effects of chemicals in humans: II. Construction of an adverse effects database for QSAR modeling. *Curr. Drug. Discov. Technol.* **1**, 243–54 (2004).
12. Hall, J.E. *Guyton and Hall textbook of medical physiology*. 13th edition. edn. (Elsevier: Philadelphia, PA, 2016).
13. Liu, T., Lin, Y., Wen, X., Jorissen, R.N. & Gilson, M.K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* **35**, D198–D201 (2007).
14. Wei, L., Altman, R.B. & Chang, J.T. Using the radial distributions of physical features to compare amino acid environments and align amino acid sequences. *Pac. Symp. Biocomput.* 465–476 (1997).
15. Miguel, A., Hsin, J., Liu, T., Tang, G., Altman, R.B. & Huang, K.C. Variations in the binding pocket of an inhibitor of the bacterial division protein FtsZ across genotypes and species. *PLoS Comput. Biol,* **11**, e1004117 (2015).
16. Brooks, P.M. & Day, R.O. COX-2 inhibitors. *Med. J. Austral.* **173**, 433–436 (2000).
17. Dwivedi, A.K., Gurjar, V., Kumar, S. & Singh, N. Molecular basis for nonspecificity of nonsteroidal anti-inflammatory drugs (NSAIDs). *Drug Discov. Today* **20**, 863–873 (2015).

18. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–D1107 (2012).
19. Bowes, J. *et al.* Reducing safety-related drug attrition: the use of in vitro pharmacological profiling. *Nature Rev. Drug Discov.* **11**, 909–922 (2012).
20. Hajduk, P.J., Huth, J.R. & Tse, C. Predicting protein druggability. *Drug Discov. Today* **10**, 1675–1682 (2005).
21. Cheng, A.C. *et al.* Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **25**, 71–75 (2007).



Supplementary information accompanies this paper on the *Clinical and Translational Science* website. (http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1752-8062)