Finding the Rules for Successful Drug Optimization

Iskander Yusof*, Falgun Shah†, Tatsu Hashimoto‡, Matthew D. Segall* and Nigel Greene†

***Optibrium Ltd., 7221 Cambridge Research Park, Beach Drive, Cambridge, CB25 9TL,UK**

† Compound Safety Prediction, Pfizer Global Research & Development, Groton, CT, USA

‡Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

Drug discovery is a process of multi-parameter optimisation, with the objective of finding compounds that achieve multiple, project-specific property criteria. These criteria are often based on the subjective opinion of the project team, but analysis of historical data can help to find the most appropriate profile. Computational 'rule induction' approaches enable an objective analysis of complex data to objectively identify interpretable, multi-parameter rules that distinguish compounds with the greatest likelihood of success for a project's objective. The importance of each property criterion highlights the most critical data that enable effective compound prioritisation decisions. We illustrate this with two applications: determining rules for simple, 'drug-like' properties; and exploring experimental target inhibition data to find rules to reduce the risk toxicity.

Drug Discovery Today (2014) 10.1016/j.drudis.2014.01.005

Introduction

A recent trend in drug discovery has been to view the process of drug design as a multi-parameter optimisation problem [1] [2], in which, from the beginning, a project team attempts to identify drug candidates that achieve an optimal balance of the biological and physicochemical properties required for a chosen therapeutic objective. By simultaneously optimising multiple properties with respect to a "profile" of several criteria, the goal is to increase the probability of quickly identifying a high quality compound over simply considering a single design parameter at a time. For example, in a project where the goal is to identify an orally dosed drug, one might define a target product profile for a lead compound similar to that shown in Table 1.

Table 1 Example property criteria for a lead compound intended for oral dosing [2].

Property	Criterion
Pharmacology	
Potency against target (K_i)	<100 nM
Selectivity against related off-targets	>100 ×
Physicochemical	
LogP	<4
Solubility	>100 μM
MW	<450 Da
ADME	
Caco-2* permeability (P_{app})	$>1\times10^{-6}$ cm/s
Intrinsic Clearance in Human Liver Microsomes (Cl_{int})	<25 μL/min/mg protein
Absence of P-glycoprotein transport (Caco2 BA:AB)	<3
Safety	
Avoid Cytochrome P450-mediated drug-drug interactions (K_i for major drug metabolising isoforms)	>1 μM
Avoid interaction with hERG potassium ion channel (IC_{50})	>10 μM
Cytotoxicity in HepG2† cells (LD_{50})	>1 mM

*Human epithelial colorectal adenocarcinoma cell line [12]
†Hepatocellular carcinoma cell line [13]

Target product profiles, such as the example in Table 1, are typically defined based on the existing body of knowledge and experience of development of orally dosed drugs. Many published examples have focused on analysis of simple 'drug-like' properties for known drugs, such as molecular weight (MW), lipophilicity (logP or logD), polar surface area (PSA), counts of hydrogen bond acceptors (HBA), donors (HBD), aromatic rings (AROM) and rotatable bonds (ROTB) [3] [4] [5] [6] [7]. These 'drug-like' profiles are sometimes based on the opinion and experience of drug discovery scientists, such as the CNS MPO score described by Wager *et al.* [8] for identifying compounds intended for central nervous system targets. In other cases, the criteria may be derived from statistical analysis of individual compound properties and combined post-hoc to form a required profile; for example, Lipinski's Rule of Five [5] and the 3/75 concept introduced by Pfizer, which relates physicochemical properties (logP > 3 and topological PSA < 75 $Å^2$) to higher likelihood of adverse outcomes in preclinical toxicology studies at a plasma C_{max} concentration less than 10µM [6].

However, as the range and volume of experimental and calculated data generated in early drug discovery increases, it is not immediately obvious how one might generalise these approaches to more complex multi-dimensional data and different drug discovery objectives, such as lowering the risk of adverse events or alternative routes of administration. A subjective approach may not necessarily yield the *optimal* property profile and one cannot use this approach to construct profiles for objectives where expert knowledge might be lacking or the range of experimental and calculated properties that could be used is very large.

Identifying an appropriate property profile is also challenging because, while it is relatively straightforward to identify trends for individual properties, a successful compound may require multiple criteria to be satisfied simultaneously. Furthermore, if some properties in the profile are highly correlated and therefore redundant, we would like to remove the redundant properties from the profile in order to focus only on the relevant subset of profile properties. This is particularly important when data is obtained experimentally, because we do not want to spend valuable time and resources generating data that adds little value to our ability to select successful compounds.

One approach to address these issues would be to automatically examine historical compound data relating a drug discovery objective to the properties of compounds and use this data to identify the key property criteria. This would enable us to process large amounts of data quickly and increase the objectivity of the analysis. However, because the available data for this analysis is unlikely to be exhaustive, it is important to also allow scientists to modify the criteria based on their knowledge of the underlying biology and chemistry. To satisfy this goal, here we will describe an approach called 'rule induction', which can analyse historical data to automatically obtain property criteria as 'rules' that are easy to interpret and modify, while still retaining objectivity as the rules are based on empirical data.

A 'rule' is a set of property criteria that can be used to identify desirable compounds. Undoubtedly, the best-known, simple example of such a rule is Lipinski's Rule of Five [9], which places bounds on the values of four physicochemical properties (MW, logP, HBA and HBD) to help medicinal chemists identify compounds with good oral absorption. A rule can also be visualised as a 'box' in property space containing significantly more desirable than undesirable compounds, as illustrated in Figure 1h. Notice that in this instance, there are multiple boxes that contain a high percentage of desirable compounds, representing different rules that select compounds in different regions of property space.

In the following section we will describe a method to identify rules from complex, multi-dimensional data, called the Patient Rule Induction Method (PRIM) [10]. We will then briefly discuss methods for application of multi-parameter rules before describing two example applications, to the derivation of rules for simple 'drug-like' properties and experimental data relating to compound toxicity.

The Patient Rule Induction Method

As discussed above, our goal of finding rules for selection of compounds with an improved chance of success can be formulated as the problem of deriving a set of boxes in property space, known as a 'box covering', from historical compound data. To do this, PRIM searches for boxes in property space over which the mean value of the objective we wish to achieve, defined as

$$\text{mean} = \frac{\text{total objective value of compounds in box}}{\text{number of compounds in box}},$$

is significantly higher than over the full property space.

For simplicity of illustration, we will describe the workings of PRIM for binary categorical objectives (good/bad), so that maximising the mean objective value is equivalent to maximising the proportion of desirable compounds. However, the method can also be used to identify desirable compounds that maximise or minimise a continuous objective value.

The first step in PRIM is to construct a single box in property space with a high proportion of desirable compounds relative to the full data set, subject to a minimum support constraint. The support is defined as the total proportion of compounds in the box relative to the full set, i.e.

$$\text{support} = \frac{\text{number of compounds in box}}{\text{number of compounds in full set}}.$$

Intuitively, we can think of the box construction strategy as a process of 'top-down peeling' followed by 'bottom-up pasting', described by the following steps and illustrated in Figure 1:

1. Start with a box in the property space covering the full set of compounds (Figure 1a)

2. Consider compressing the box at each 'face' in turn; this is called 'peeling'

3. Remove the 'face' that results in the largest increase in the proportion of desirable compounds (Figure 1b). The amount of data removed at each peel is controlled by a 'peeling fraction', which is typically small, e.g. 5%. The small proportion of data removed at each step gives rise to the term *patient* rule induction, in contrast with traditional recursive partitioning, which removes 50% of the data with each step.

4. Repeat steps 1 and 2 until a small box is obtained that contains a high percentage of desirable compounds, subject to a minimum support constraint to avoid overtraining (Figure 1c)

5. The peeling process is not guaranteed to produce the optimal box, as each peel is made based only on the information available to it at the time; hence the peeling process is reversed by repeatedly 'expanding' each box face in turn as long as the proportion of desirable compounds can be increased. This is called 'pasting' (Figure 1d).

6. Repeat step 5 until the proportion of desirable compounds in the box can no longer be increased (Figure 1e)

7. This peeling-and-pasting process results in a 'peeling sequence' of boxes, ranging from those with a high proportion of desirable compounds and low support to those with a high support and (relatively) low proportion of desirable compounds

8. The box chosen from the peeling sequence is the one that maximises the proportion of desirable compounds selected from the validation set, subject to a user-specified minimum support (Figure 1f)

This gives us a single box, corresponding to a single rule, but it may be that there are additional boxes containing a high proportion of desirable compounds that can be discovered by "looking elsewhere" in property space. To search for another such box, the compounds contained by the first box obtained can be removed (Figure 1g) and Steps 1-8 above repeated. This process is, in turn, repeated until no further boxes can be found. The final result will be a 'covering' of boxes that collectively describe the regions of property space where the proportion of desirable compounds is high (Figure 1h).

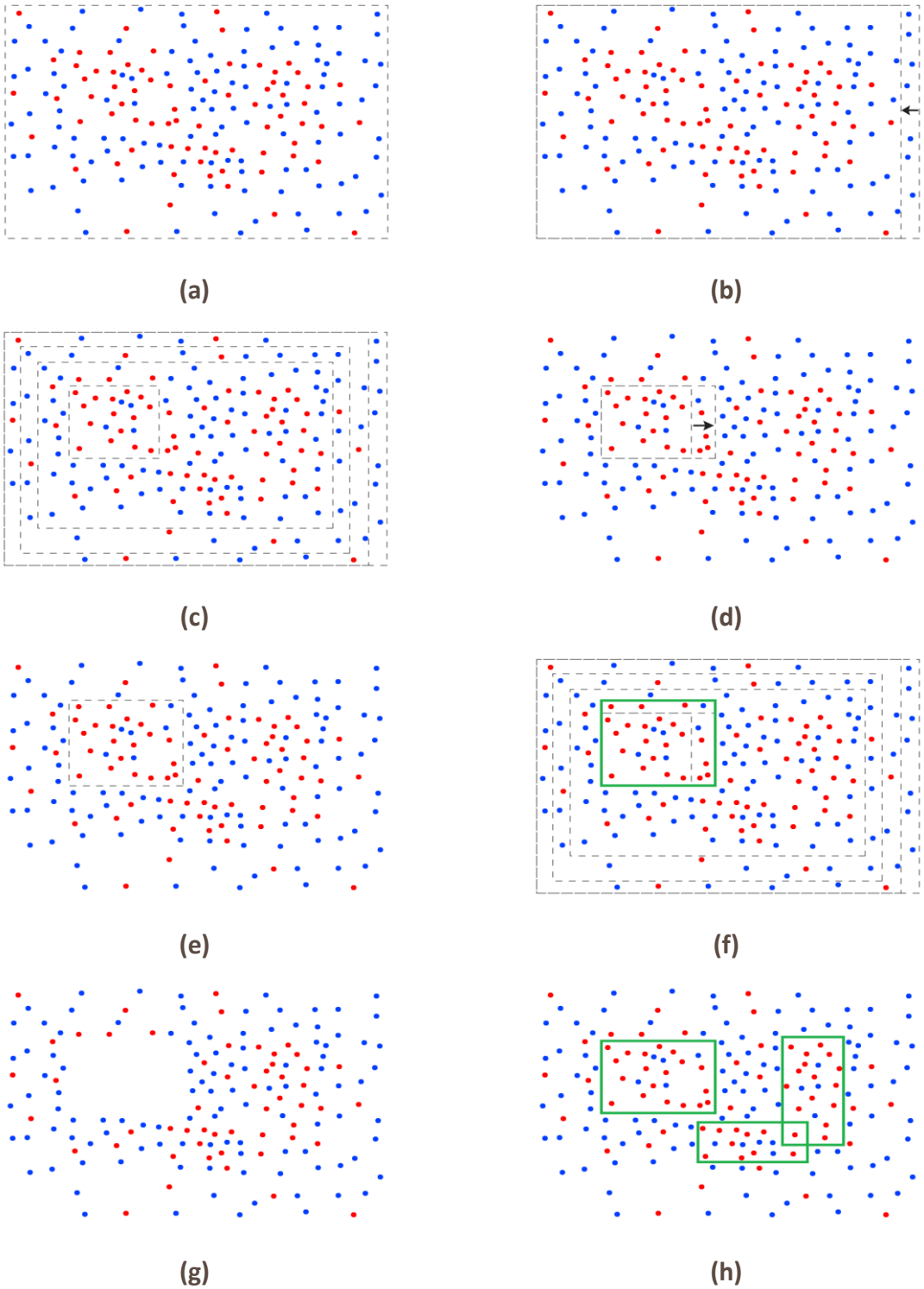


Figure 1. An illustration of the PRIM algorithm to identify boxes in a two-dimensional property space corresponding to rules for selection of compounds with a high chance of achieving a desired objective. The corresponding steps in the process are described in the text. The points in each plot correspond to compounds with existing data, plotted in the property space; desirable compounds are highlighted in red, undesirable compounds in blue. Boxes shown as dashed lines correspond to potential rules, and those highlighted in green correspond to those selected by the algorithm.

We now have a series of boxes, each corresponding to an individual rule, that will increase the probability of identifying desirable compounds. However, we might also wish to know which of the property criteria defining a rule are most important to determining the rule's ability to distinguish 'good' from 'bad' compounds.

This importance of an individual criterion is determined in a probabilistic manner, by considering the false negative rate of the criterion; the importance of each property criterion is defined to be 1 minus the false negative rate of the criterion, as computed over the training set. In other words, the most important property criteria are those where there is a very low probability of finding a desirable compound that fails to meet the criterion. The resultant weighting for each criterion will range from 0 to 1, with 0 indicating no importance and 1 being the maximum importance value. In addition to helping the chemist decide whether any properties can be compromised without significantly affecting the odds of satisfying the objective, these weightings provide an indication as to what (experimental or calculated) property data should be generated to aid compound prioritisation decisions.

Applying Multi-Parametric Rules

Whether a rule is derived based on scientists' experience, analysis of historical data or a combination thereof, there are a number of approaches with which to apply the rule to the selection of compounds [2].

Probably the most common approach is to apply the property criteria corresponding to a rule as filters, rejecting compounds that do not meet a defined cut-off in each parameter. While simple to apply and interpret, this approach has significant drawbacks. In particular, hard cut-offs draw inappropriately harsh distinctions between similar compounds; for example, a common criterion for logP is that it should be less than 5, but does a compound with a logP of 5.1 carry a significantly greater risk than another with logP of 4.9? This issue is further compounded by the fact that many of these properties have significant uncertainty in the reported values. Again, considering the example of logP, a predictive model may have an uncertainty of 0.5 log units, meaning that the two hypothetical compounds discussed previously may not even be confidently distinguishable on the basis of these data.

An alternative approach is to define each property criteria as 'desirability function' that defines the desirability of the potential values of a property on a scale of 0 to 1. A desirability function can take any shape, allowing the impact of a poor outcome to be weighted appropriately against other factors and 'softening' hard cut-offs. The desirabilities of multiple properties can then be combined to obtain an overall "score" for the compound reflecting the overall balance of properties. An example of this approach is the quantitative estimate of drug-likeness (QED) method as proposed by Bickerton *et al.* [11].

The uncertainty in the underlying data, due to experimental variability or statistical error, can be explicitly taken into account using a method such as probabilistic scoring [12]. This method builds on the foundation of desirability functions to calculate a score reflecting the likelihood of success of each compound against the property profile, taking into account both the importance of each criterion to the overall project objective and the uncertainty in each data point. The uncertainty in the overall score can also be estimated, making it clear when compounds can be confidently distinguished and avoiding missed opportunities due to inappropriately rejecting compounds based on uncertain data. This is illustrated in Figure 2, which shows an example of a scoring profile and resulting output of probabilistic scoring. In the examples below, the rules derived by rule induction were applied using the probabilistic scoring method implemented in StarDrop version 5.5 [13].

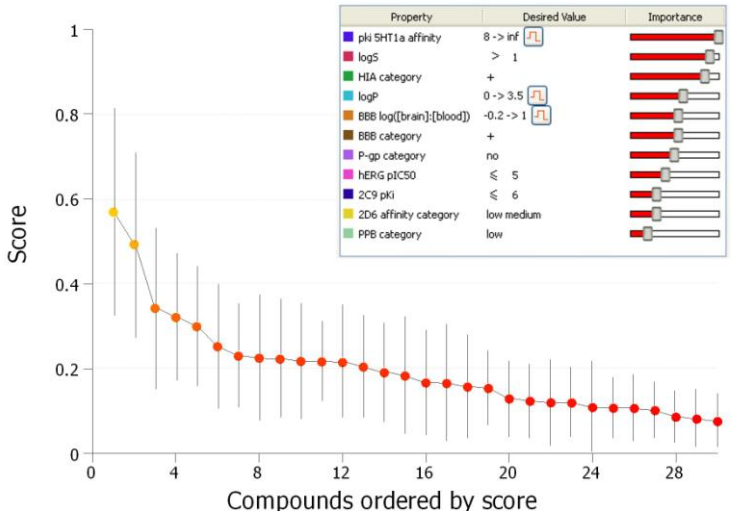


Figure 2. An example output from probabilistic scoring for 30 compounds. The compounds are ordered from left to right along the x-axis in order of their score and overall score for each compound is plotted on the y-axis. The overall uncertainty in each score (one standard deviation), due to the uncertainty in the underlying data, is shown by error bars around the corresponding point. From this it can be seen that approximately the bottom 50% of compounds may be confidently rejected, as their error bars do not overlap with that of the top-scoring compound.

The scores have been calculated against the inset scoring profile, showing the property criteria and importance of each criterion to the overall project objective. Underlying each criterion is a desirability function.

Example 1: Identifying Drug-like Compounds

As discussed above, many measures of 'drug-likeness' have been discussed in the literature, relating easily calculated molecular properties to outcomes such as oral activity or 'developability' [9] [5] [4] [6] [11]. Many of these take the form of simple rules, but Bickerton *et al.* describes a quantitative metric, QED, based on a combination of the outputs of desirability functions for logP, HBA, HBD, PSA, ROTB, AROM and the number of alerts for undesirable functionalities (ALERT) [11]. Each desirability function corresponds to a single molecular property, and is derived empirically by fitting to this property's distribution over a set of 771 approved oral drugs. To compute the QED score for an individual compound, these desirability functions are combined by taking the geometric mean of all eight desirability scores, giving an overall QED score ranging from 0 (all properties are completely undesirable) to 1 (all properties are ideal).

An issue with this approach is that the QED score for a compound is based solely on the property distributions of a set of approved oral drugs; it does not take into account whether these distributions can *differentiate* the drugs from the 'non-drugs', i.e. the other compounds that might be synthesised. For this reason, an alternative approach, the Relative Drug Likelihood metric (RDL) [14], defines a compound's desirability score for a property to be the relative probability of obtaining this compound's property value if it is a drug versus a 'non-drug'.

However, both of these approaches only consider the effect of one property at a time on the drug classification and combine these properties post-hoc. Conversely, PRIM considers all property criteria simultaneously to find those criteria that, in combination, distinguish drugs from non-drugs. Furthermore, PRIM will also tell us whether any of these eight properties are redundant for the objective of classifying a compound as a drug or non-drug.

Figure 3 shows the rules obtained by PRIM when applied to a data set comprising 771 'positive' oral drugs and 1,000 'negative' non-drug compounds randomly selected from ChEMBL for the objective of identifying oral drugs. The data set was randomly split into training and validation sets containing 70% and 30% of the compounds respectively, and we generated rules using minimum support values of 20% and 50%. Notice that the rules obtained only contain criteria for a subset of the eight properties used to train the algorithm, indicating that the excluded properties do not impart a significant amount of extra information about the objective compared with the subset of properties chosen by PRIM.

One common way to assess the performance of classifiers is a Receiver Operating Characteristic (ROC) curve. Figure 4 shows the performance of QED, RDL, and the rules generated by PRIM on the task of differentiating an independent test set of 247 oral drugs from 1000 non-drugs randomly selected from ChEMBL (different from those used to find the rules). Although the PRIM rules only specify criteria for a subset of the original eight properties, they are able to match the performance of RDL on this benchmark. Note also that QED performs poorly in this instance with an AUC of just 0.52, showing that the choice of 'negative' set has a substantial impact on the effectiveness of this metric.

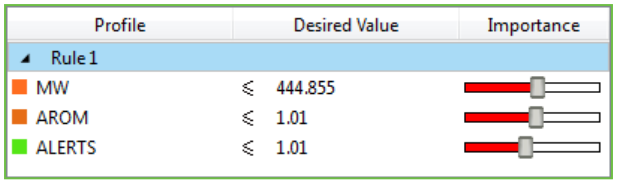
Example 2: Toxicity Classification

We also applied PRIM to search for rules that help distinguish compounds exhibiting *in vivo* toxicity from those that do not have significantly increased risk, based on experimentally measured *in vitro* data. In this example, we have explored data sets of known drugs to their cardiotoxic and hepatotoxic potential in clinic using set of biochemical assays. These drugs were profiled in CEREP Bioprint® assays panel (see http://www.cerep.fr/Cerep/Users/pages/productsservices/bioprintservices.asp), which offers biochemical assay against 185 targets including GPCR, kinase, nuclear hormone receptors and Cytochrome P450s, etc., and assess the extent of off-target pharmacology of these compounds. These biochemical assays were run at single concentration of 10 µM. A full list of the targets is included in the supplementary information (Table S1). A reporting odds ratio (ROR) was used to detect a signal of potential drug-adverse event association using information from the FDA Adverse Event Reporting system (FAERS, formerly AERS) database (see http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm), which contains voluntary reports of adverse events submitted to the FDA by healthcare practitioners, manufacturers and consumers. The ROR signals for cardiotoxicity and hepatotoxicity were calculated for these drugs as reported elsewhere [15]. A ROR signal cut-off of 2.5 or above at the System Organ Class (SOC) level in the MedDRA Ontology was used to classify compounds as having cardiac or hepatic risks in the clinic, whereas a ROR signal of less than 2.5 was used to classify compounds as having no cardio or hepatotoxicity. Tables S2 and S3 list these compounds and their toxicity classification based on the ROR signal cut-off. We split each of the two data sets into independent training, validation, and test sets comprising 70%, 15%, and 15% of the full set respectively.

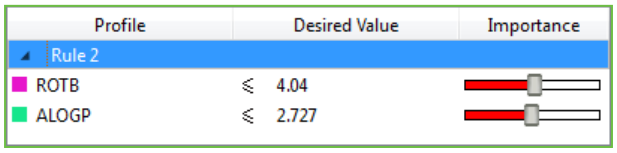
The first data set consisted of 474 known drugs, 408 of which were labelled as 'cardiotoxic' and 66 as 'non-cardiotoxic' based on the ROR signal cut-off. It is worth noting that many of the drugs classified as cardiotoxic here are in fact used in the treatment of cardiovascular diseases and so their 'toxicity' may result from either the underlying disease state or from the intended pharmacology of the drug in question. PRIM generated rules comprising property criteria for increasing the probability of selecting non-cardiotoxic compounds from the training set, which are validated using compounds in the test and validation sets.

Figure 3c shows the rule obtained using a minimum support value of 8%. It is worthwhile to mention that the algorithm has only used the three most predictive properties (out of a total of 185 properties) in order to prevent overtraining. The rule exhibits a large mean improvement of 419% over the test set, and the ROC curve (Figure 4b) generated from the test set compounds shows that the rule performs well at selecting non-cardiotoxic compounds. 5 of the 6 test set compounds selected by the rule are non-cardiotoxic, whereas only 13 of 81 compounds in the full test set are non-cardiotoxic. This means that over 83% of the compounds selected by the rule are non-cardiotoxic, so the rule offers a substantial improvement over chance, as we would expect roughly 16% of the selected compounds to be non-cardiotoxic if we randomly guessed which compounds were non-cardiotoxic. Furthermore, of the 20 test set compounds that fail every criterion in the rule, 19 are cardiotoxic, implying that any compound failing all the criteria comprising the rule has a very high chance of being cardiotoxic.

The specific values of the three property criteria identified are features of the measurements for specific compounds in the data set and could be manually rounded to more easily recognised values. However, it is clear that they essentially correspond to absence of inhibition of the histamine 2 (H2), serotonin 5-Hydroxytryptamine (5-HT_{1A}) and adenosine 1 (A1) receptors, which is biologically plausible, because interaction with these receptors have been previously associated with cardiotoxicity. For example, activation of 5-HT_{1A} are known to cause decrease in blood pressure and heart rate via modulation of sympathetic nerve activity [16] [17]; Likewise, stimulation of adenosine receptor cause bradycardia and hypotension [18]and the activation of H2 receptor are reported to cause vasorelaxation as reported by Jansen-Olesen *et al.* [19].

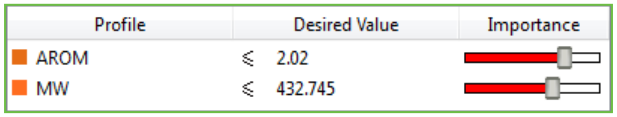


Set	Mean Improvement (%)	Support (%)
Train	60	22
Val	57	24



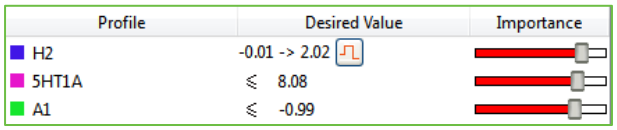
Set	Mean Improvement (%)	Support (%)
Train	51	23
Val	46	23

(a)



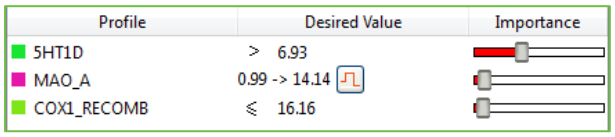
Set	Mean Improvement (%)	Support (%)
Train	35	57
Val	35	58

(b)



Set	Mean Improvement (%)	Support (%)
Train	233	9.3
Val	173	10
Test	419	7.4

(c)



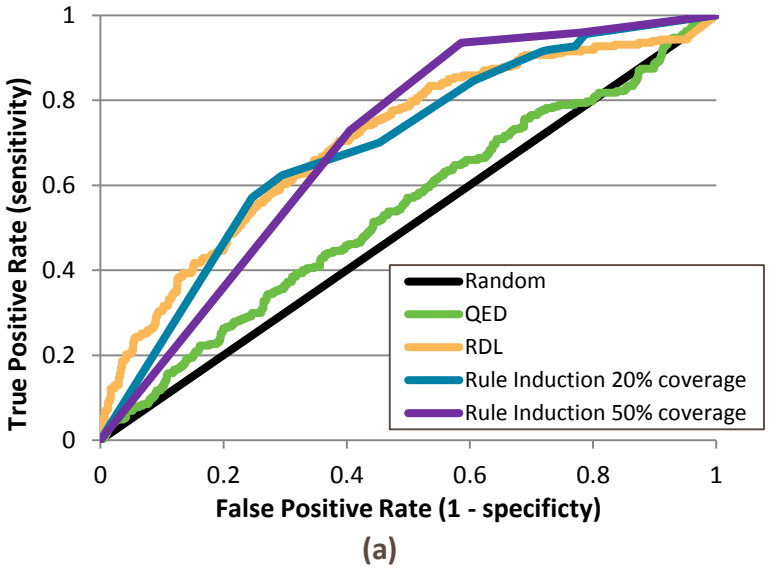
Set	Mean Improvement (%)	Support (%)
Train	51	12
Val	56	14
Test	39	11

(d)

Figure 3. Examples of multi-parameter scoring profiles derived using PRIM. (a) and (b) show rules for identifying drug-like compounds obtained by PRIM when applied to a data set comprising 771 "positive" oral drugs and 1,000 "negative" non-drug compounds randomly selected from ChEMBL. The corresponding predictive performance of these rules over the training and validation sets is also shown, for rules were generated with a minimum support of 20% (a) and 50% (b).

(c) a rule for identifying non-cardiotoxic compounds obtained by PRIM when applied to a data set consisting of 408 cardiotoxic and 66 non-cardiotoxic compounds. The corresponding predictive performance of this rule over the training, validation, and test sets is also shown. The criteria correspond to percentage inhibition of the histamine 2 (H2), serotonin receptor 5-HT_{1A} and adenosine 1 (A1) receptors.

(d) a rule for identifying non-hepatotoxic compounds obtained by PRIM when applied to a data set consisting of 168 hepatotoxic and 302 non-hepatotoxic compounds. The corresponding predictive performance of this rule over the training, validation, and test sets is also shown. The criteria correspond to percentage inhibition of the serotonin 5-HT_{1D} receptor and monoamine oxidase A (MAO_A) and cyclooxygenase 1 (COX1) enzymes.



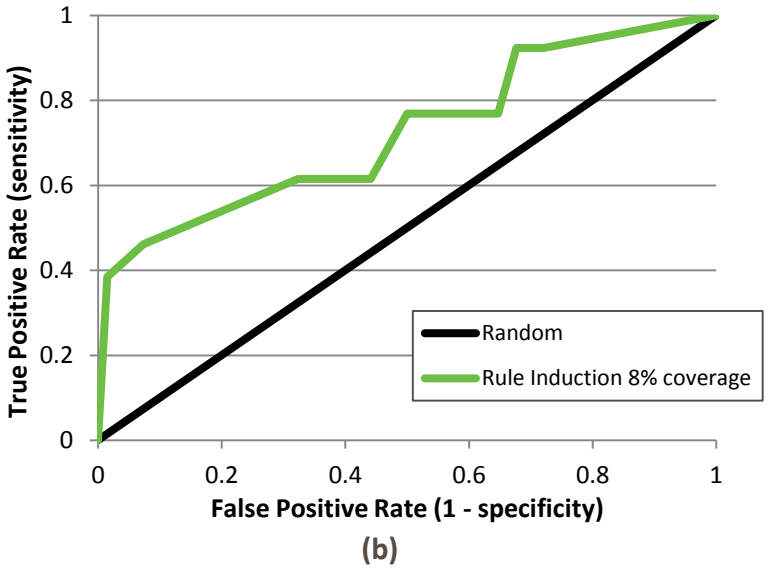


Figure 4 ROC plots of the true positive rate (TPR (sensitivity)) against the false positive rate (FPR (1 - specificity)) for the classification of compounds. A perfect classifier would be represented by the point in the top left and a performance below the identity line indicates worse performance than a random classification. A greater area under the curve (AUC) for a classifier indicates higher performance.

(a) ROC plot for classification of compounds as orally absorbed drugs or otherwise using RDL, QED, and PRIM with a minimum support of 20% and 50%. In this case, a set of 247 orally administered drugs was differentiated from 1,000 randomly selected compounds from ChEMBL; the AUC for QED is 0.52, RDL is 0.70, and PRIM is 0.69 for a minimum support of 20% and 0.70 for a minimum support of 50%.

(b) Classification of compounds as non-cardiotoxic or otherwise using the rule in Figure 3(c) derived with PRIM. Here, a set of 66 non-cardiotoxic compounds was differentiated from 408 cardiotoxic compounds. The AUC in this case is 0.72.

The second data set contained 470 compounds, 302 of which were labelled as 'hepatotoxic' and 168 as 'non-hepatotoxic'. Here we searched for rules to increase the probability of selecting non-hepatotoxic compounds based on a minimum support value of 10%.

Figure 3d shows the rule obtained for selection of non-hepatotoxic compounds. The rule shows a reasonable mean improvement of 39% over the test set, with 9 of the 10 test set compounds being non-hepatotoxic versus 51 non-hepatotoxic compounds out of 80 in the full test set. However, the property criteria themselves do not appear to be biologically relevant. The rule relates to binding to the 5-hydroxytryptamine 1D (5-HT_{1D})

receptor, mono-amine oxidase A (MAO-A) and cyclooxygenase 1 (COX1) enzymes. However, the criterion for 5-HT_{1D} inhibition suggest that an *increased* inhibition of this enzyme reduces the risk of hepatotoxicity, while the criterion for inhibition of MAO-A suggest a narrow range of inhibition reduces hepatotoxicity risk; both of which are implausible. These statistically significant correlations may arise due to chance in a relatively small and noisy data with many properties or may be due to correlation of a property with another causative relationship. This demonstrates the advantage of outputting rules as interpretable property criteria over a "black-box" classifier; even if a rule appears to offer good predictive performance, we may still wish to discard or modify it based on an expert's understanding of the specific property criteria comprising the rule. In this case, a plausible rule has not been found because the large majority of the targets for which data are present in the data set are not known to relate with hepatotoxicity. In the few examples of targets that are known to correlate with this toxic outcome, such as $PPAR_{\gamma}$ [20] [21], there are a statistically insignificant number of inhibitors in the data set and hence no correlation could be found. This reinforces the point that any method will be limited by the quality of data available.

Conclusions

Making progress in drug discovery relies on making good decisions based on complex, multi-parameter data. MPO methods provide reasonable approaches to assess compound data for multiple properties against a profile of criteria, to identify the compounds most likely to achieve a drug discovery project's objective. However, a question often remains regarding the most appropriate property profile to use for a specific project's therapeutic objectives.

In this paper, we have discussed an approach based on PRIM and shown that it can generate multi-parameter rules for identifying desirable compounds from both calculated and experimental data. These rules are highly interpretable and easy to modify based on expert opinion, while still offering good predictive performance on unseen data. Furthermore, on data sets with large numbers of properties, we have demonstrated a method for determining the importance of each property contributing to the objective, with only properties relevant to the objective forming part of the final rule(s). This will allow experimental resources to be prioritised appropriately, with data for highly important properties being more valuable to generate than data for other, less-relevant properties.

The utility of PRIM was demonstrated by identifying rules in two example applications, one for deriving rules for 'drug-like' properties and the other for distinguishing between toxic and non-toxic compounds. Using an MPO method, such as probabilistic scoring, these rules can then be applied prospectively to new data sets to select and prioritise compounds.

One limitation to the use of these rule induction methods is that, like any conventional modelling method, they require a statistically significant number of data points from which to derive multi-parameter rules. In many cases, the number of compounds for which results are available for complex endpoints, such as clinical success, is limited. The data set may also be highly biased toward one outcome. However, as illustrated by the cardiotoxicity example above, rule induction methods have been shown to be robust to bias in the data set, and often perform better when the minority of the results correspond to the desired outcome. This is because rule induction attempts to identify rules that significantly increase the mean value of the objective and contrasts with most conventional machine learning algorithms, which typically perform better for classification of the majority class. Furthermore, where necessary complex objectives may be broken down into sub-objectives for which sufficient data are available, e.g. bioavailability, exposure, safety, etc., and the resulting property rules combined to identify those compounds which are most likely to satisfy all of the sub-objectives.

Supplementary Information

Table S1 lists the targets for which inhibition data are included in the data sets used to identify rules for hepatotoxicity and cardiotoxicity.

Table S2 lists the drugs used to derive rules for reduced cardiotoxicity risk together with their cardiotoxic classification based on a ROR signal cut-off of 2.5.

Table S3 lists the drugs used to derive rules for reduced hepatotoxicity risk together with their hepatotoxic classification based on a ROR signal cut-off of 2.5.

The positive and negative data sets used to derive and test the rules for 'drug-like' properties for oral drugs are provided.

References

1 Ekins S, Boulanger B, Swaan P, Hupcey M. Towards a new age of virtual ADME/TOX and multidimensional drug discovery. J. Comp. Aided Mol. Design. 2001;16:381-401.

2 Segall MD. Multi-Parameter Optimization: Identifying high quality compounds with a balance of properties. Curr. Pharm. Des. 2012;18(9):1292-1310.

3 Leeson PD, Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. Nat. Rev. Drug Discov. 2007;6(11):881-890.

4 Ritchie TJ, Macdonald SJF. The impact of aromatic ring count on compound developability – are too many aromatic rings a liability in drug design? Drug Discov. Today. 2009;14(21/22):1011-1020.

5 Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. J. Med. Chem. 2002;45(12):2615-2623.

6 Hughes JD, Blagg J, Price DA, Bailey S, Decrescenzo GA, Devraj RV, Ellsworth E, Fobian YM, Gibbs ME, Gilles RW, et al. Physicochemical drug properties associated with in vivo toxicological outcomes. Bioorg. Med. Chem. Lett. 2008;18(17):4872-4875.

7 Lovering F, Bikker J, Humblet C. Escape from flatland: increasing saturation as an approach to improving clinical success. J. Med. Chem. 2009;52(21):6752-6756.

8 Wager TT, Hour X,VPR, Villalobos A. Moving beyond rules: The development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. ACS Chem. Neurosci. 2010;1(6):435-449.

9 Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 1997;23:3-25.

10 Friedman J, Fisher N. Bump hunting in high-dimensional data. Statistics and Computing. 1999;9:123-43.

11 Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL. Quantifying the chemical beauty of drugs. Nature Chemistry. 2012;4(2):90-98.

12 Segall MD, Champness E, Obrezanova O, Leeding C. Beyond Profiling: Using ADMET models to guide decisions. Chemistry and Biodiversity. 2009 2144-2151.

13 Optibrium. [Internet]. [cited 2013 May 28]. Available from: http://www.optibrium.com/stardrop.

14 Yusof I, Segall MD. Considering the impact drug-like properties have on the chance of success. Drug Discovery Today. 2013;18(13-14):659-666.

15 Nadanaciva S, Aleo MD, Strock CJ, Stedman DB, Wang H, Will Y. Toxicity assessments of nonsteroidal anti-inflammatory drugs in isolated mitochondria, rat hepatocytes, and zebrafish show good concordance across chemical classes. Toxicol. Appl. Pharmacol. 2013;272(2):272-280.

16 Dabiré H. Central 5-hydroxytryptamine (5-HT) receptors in blood pressure regulation. Therapie. 1991;46(6):421-429.

17 Ramage AG. Influence of 5-HT1A receptor agonists on sympathetic and parasympathetic nerve activity. J. Cardiovasc. Pharmacol. 1990;15(Suppl 7):S75-85.

18 Bonizzoni E, Milani S, Ongini E, Casati C, Monopoli A. Modeling hemodynamic profiles by telemetry in the rat. A study with A1 and A2a adenosine agonists. Hypertension. 1995;25(4 Pt 1):564-569.

19 Jansen-Olesen I OACL, Strunk S, Lassen LH, Olesen J, Mortensen A, Engel U, L. E. Role of endothelium and nitric oxide in histamine-induced responses in human cranial arteries and detection of mRNA encoding H1- and H2-receptors by RT-PCR. Br. J. Pharmacol. 1997;121(1):41-48.

20 Panasyuk G, Espeillac C, Chauvin C, Pradelli LA, Horie Y, Suzuki A, Annicotte JS, Fajas L, Foretz M, Verdeguer F, et al. PPARγ contributes to PKM2 and HK2 expression in fatty liver. Nat. Commun. 2012;3(672):14.

21 Rogue A, Lambert C, Jossé R, Antherieu S, Spire C, Claude N, Guillouzo A. Comparative gene expression profiles induced by PPARγ and PPARα/γ agonists in human hepatocytes. PLoS One. 2011;6(4):e18816.