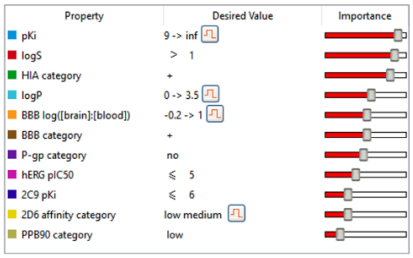
Avoiding Missed Opportunities by Analysing the Sensitivity of our Decisions

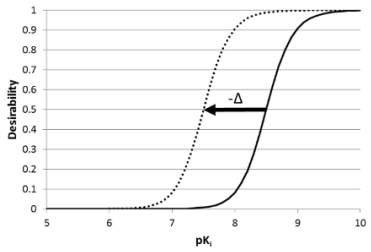
Matthew D. Segall, Iskander Yusof and Edmund J. Champness

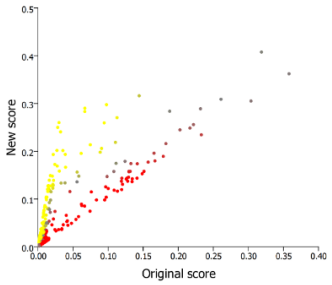
Optibrium Ltd., 7221 Cambridge Research Park, Beach Drive, Cambridge, CB25 9TL, UK

Abstract

Drug discovery is a multi-parameter optimisation process, in which the goal of a project is to identify compounds that meet multiple property criteria required to achieve a therapeutic objective. However, having chosen a profile of property criteria, their impact on the decisions made regarding progression of compounds or chemical series should be carefully considered. In some cases, the decision will be very sensitive to a specific property criterion and such a criterion may be artificially distorting the direction of the project; any uncertainty in the 'correct' value or the importance of this criterion may lead to valuable opportunities being missed. In this paper, we will describe a method for analysing the sensitivity of the prioritisation of compounds to a multi-parameter profile of property criteria. We show how the results can be easily interpreted and illustrate how this analysis can highlight new avenues for exploration.







7221 Cambridge Research Park
Beach Drive, Cambridge
CB25 9TL, UK

Tel: +44 1223 815900
Fax: +44 1223 815907

Email: info@optibrium.com
Website: www.optibrium.com

Introduction

Medicinal chemistry is a delicate balancing act, requiring many compound properties to be simultaneously optimised to achieve the therapeutic objectives of a drug discovery project. A safe and efficacious drug that reaches the market must have sufficient potency against its therapeutic target(s), appropriate absorption, distribution, metabolism and excretion (ADME) properties and avoid interactions with off-targets or non-specific toxicities at therapeutic concentrations to achieve an acceptable safety window.

In order to guide this multi-parameter optimisation (MPO) process, it is common to define a set of property criteria or a 'target profile' that compounds must achieve for progression at different stages, e.g. hit selection, hit-to-lead and candidate selection. These are often defined by a multi-disciplinary project team, based on the experience of medicinal chemistry, pharmacology, drug metabolism and pharmacokinetics (DMPK) and toxicology team members, sometimes supported by statistical analysis of other compounds previously progressed for similar objectives [1] [2].

However, having chosen a profile of property criteria, we should consider the impact that this choice will have on the decisions made, i.e. the compounds or chemical series chosen for progression. In some cases, the choice of compounds will be very *sensitive* to a specific property criterion or the importance given to it. In these cases, that criterion may be artificially distorting the direction of the project; if we are not very confident that the 'right' criterion has been used, this may lead to valuable opportunities being missed.

As an analogy, consider another decision we may make in our everyday lives; choosing the 'perfect' hotel for a well-deserved vacation. We might look on Trip Advisor and consider many criteria including location, reviews of characteristics such as cleanliness of accommodation, proximity to an attraction, family friendliness and others. Maybe we decide that the hotel must be within half a mile of the beach and, to narrow down the choices, we filter out all of the hotels that are further away. But, what if an otherwise perfect hotel was 900 yards from the beach? This arbitrary decision may mean that we miss the vacation of a lifetime!

In the context of drug discovery, consider a progression criterion for potency that specifies that the IC_{50} must be less than 10nM. But, what if the most active member of a chemical series with good ADME and safety characteristics has a potency of 50 nM; would it make sense to reject this series? In a simple case such as this, it may be possible to spot this exception, but with the increasing complexity and diversity of the data used in early drug discovery, these sensitivities may not always be apparent. In addition, as time progresses it can be difficult to remember the details of chemical series explored earlier in a project and consideration of these sensitivities can reveal alternative directions or backup series, should a project reach an insurmountable issue with its primary series.

In another scenario, it is not uncommon to have disagreements within a project team regarding the most appropriate property criteria and their importance to the overall decision. For example, one member of a team may insist that the aqueous solubility of a compound must be at least 100 µM to consider for progression, while another may feel that this is too strict, suggesting that any downstream solubility issues could be addressed by appropriate formulation. In this scenario, the question is, "Does it matter?" If both criteria lead the project to select the same compounds, then this debate is not relevant to the decision being made. On the other hand, if the choice of solubility criterion will dramatically affect the direction of the project, the debate must be had in earnest. Alternatively, it may be a good strategy to sample some of the compounds selected by each criterion for progression and gather some downstream data before making a final decision about which criterion is the most appropriate.

Answers to questions such as these can be addressed by sensitivity analysis, to rigorously consider the impact of small changes in property criteria and their importance in the context of a multi-parameter profile. In the paper, we will describe a method for sensitivity analysis that generates easily interpretable results, to help to guide critical compound selection decisions in drug discovery, avoiding missed opportunities and revealing potentially valuable new avenues for exploration.

In the next section we will give a brief overview of approaches to multi-parameter optimisation as a foundation for a description of methods for sensitivity analysis. We will then discuss the application of these methods using some illustrative applications before drawing conclusions.

Methods

Introduction to Multi-Parameter Optimisation

Methods for prioritising solutions (compounds, in the context of drug discovery) are commonly described using a range of terms including MPO, multi-objective optimisation (MOOP), and multi-criteria decision-making (MCDM). For simplicity, we will refer to these as MPO. Comprehensive reviews of MPO methods applied in drug discovery can be found in references [3] [4].

Probably the simplest approach for MPO is the use of filters, removing compounds that do not meet each of a series of property criteria in the hope that a 'perfect' compound will emerge that passes all of the criteria. However, particularly early on in a drug discovery project, it is unlikely that any compound under consideration will meet all of the criteria and, in practice, it is usually necessary to compromise in order to make progress. In such a scenario, the importance or weight of each criterion should be considered. Furthermore, filters are often unnecessarily harsh, making inappropriate distinctions between compounds. For example, a popular selection criterion based on molecular weight (MW) is that a compound should be less than 500 Da to be 'drug like'. However, does a compound with a MW of 499 Da really have a better chance of success than another compound with MW of 501 Da? The correlation between MW and the success of compounds does not support this distinction [5]. This problem is compounded by the fact that the data with which decisions are made often have significant uncertainties due to experimental variability or statistical error [6]. For example, the standard deviation in a measured IC_{50} may be a factor of 3 (0.5 log units), so in this case a compound with a measured IC_{50} of 125 nM has, approximately, a 40% chance of meeting a criterion of <100 nM (assuming a Normally distributed error). Furthermore, the individual uncertainties in each filter combine; for example applying 5 filters, each of which has an accuracy of 80%, means that the chance of a perfect compound passing all 5 filters is only 33%, i.e. the sequence of filters is twice as likely to reject a perfect compound as to pass it. Therefore, despite its apparent simplicity, filtering should be treated with caution.

One approach to avoiding the artificial harshness of filters is to use *desirability functions* [7] to map the value of each property onto a scale of desirability from 0 (unacceptable) to 1 (ideal). This enables more subtle distinctions to be made between similar property values. Some simple examples of desirability functions are shown in Figure 1. The desirabilities of individual properties can be combined to calculate an overall *desirability index*, or score, that can be used to rank compounds according to their performance against multiple criteria simultaneously. There are two common approaches to combining the desirabilities of individual properties to calculate an overall desirability index *D*:

Additive: $D(x_1, x_2, \dots, x_N) = \sum_{i=1}^{N} c_i d_i(x_i)$;

Multiplicative: $D(x_1, x_2, \dots, x_N) = \prod_{i=1}^{N} d_i(x_i)^{c_i}$;

where x_i are the values of *N* properties, d_i are the desirability functions for the corresponding properties and c_i are optional coefficients that can be used to define the weight of each individual property in the profile. The scores may also be normalised for the number of properties being considered by taking the arithmetic or geometric mean in the case of the additive or multiplicative approaches respectively.

An additive approach has a disadvantage that the impact of any single property on the overall score will be limited; a poor result for a single property is not sufficient to 'kill' a compound, which may be appropriate in some cases. For example, a completely inactive compound would not be interesting, even if it had perfect ADME and physicochemical properties. Furthermore, unless it is appropriately normalised, an additive score will increase with the number of properties for which data has been generated, which may inappropriately bias decisions towards compounds which have progressed further and hence been studied in more assays. In contrast, with a multiplicative approach, critical properties can be defined, such that a poor result in any of these would be sufficient to 'kill' a compound. This provides greater flexibility in the definition of the required property profile, however there is a danger that a multiplicative approach could be overly harsh. Where a poor result in one property could be mitigated by good outcomes for others, the desirability functions should be carefully chosen to avoid this. For example, for a CNS indication, limited blood-brain-barrier penetration could be offset by high potency and low protein binding and therefore it may be inappropriate to assign a desirability close to zero for moderate blood-brain barrier penetration.

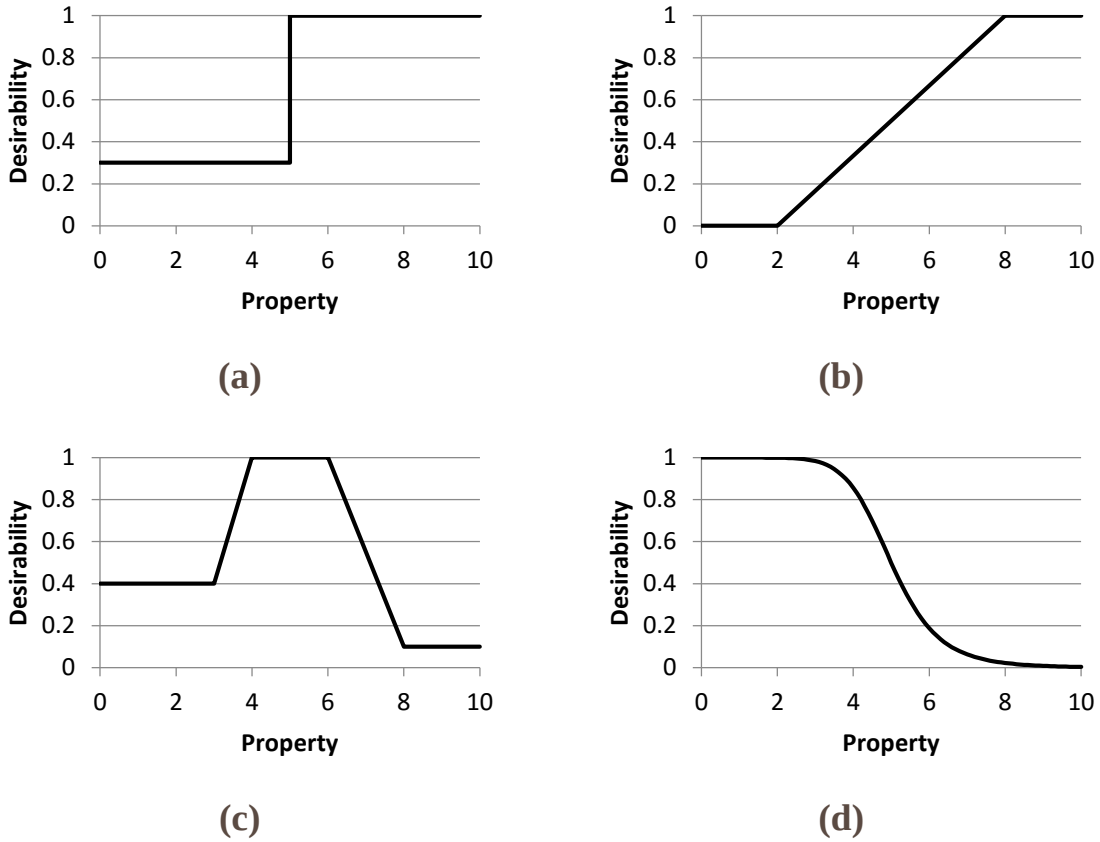


Figure 1. Example desirability functions. (a) represents a hard cut-off, but not a filter, because property values below 5 are less desirable, but will not be rejected outright; (b) defines a linearly increasing desirability between 2 and 8 with values below 2 being unacceptable and those above 8 being ideal; (c) defines an ideal property range of 4-6 with values exceeding the upper limit being less desirable than those below the lower limit. The boundaries of this range are not hard cut-offs, but are 'softened' because the desirability above and below the ideal range decreases linearly; (d) s an example of a non-linear sigmoidally decreasing desirability function with an inflection at a property value of 5.

Although not the first examples of the application of desirability functions in drug discovery, their use has been popularised by two recently published methods: the quantitative estimate of drug-likeness (QED) [8] and CNS MPO [9]. The former was developed for the selection of compounds with an improved chance of success as an oral drug, the latter for selection of compounds for central nervous system (CNS) indications. However, these methods should be used with caution because, in practice, they do not perform much better than random selection for these objectives [10] [11]. This is because the simple compound properties on which they are based, such as MW, lipophilicity and polar surface area, do not have a strong correlation with the ultimate *in vivo* behaviour of a compound. However, desirability functions can be applied to compound data, such as the results of *in vitro* assays or predictions from predictive models of potency, physicochemical and ADME properties, where there is a stronger correlation with the *in vivo* outcome.

While desirability functions move away from the 'black and white' distinctions of filters, it is also valuable to explicitly consider the uncertainties in the underlying data, so that it is clear when compounds can be confidently distinguished or, alternatively, when the data do not support such a conclusion. This is important because inappropriately rejecting a compound or series due to an uncertain measurement or prediction may artificially narrow the scope of exploration in a project and lead to missed opportunities. Given that high quality compounds are rare, the cost of missing an opportunity is likely to be high.

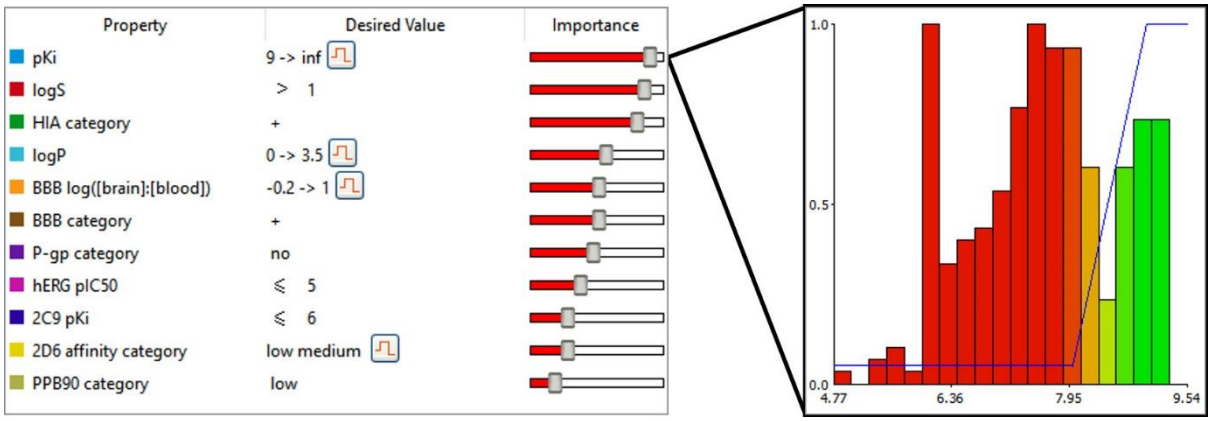


Figure 2. An example of a profile of property criteria suitable for identifying compounds that are potent (high pK_i) and with suitable physicochemical and ADME properties for oral dosing against a target in the central nervous system. Underlying each of the criteria is desirability function, as illustrated for the pK_i. The histogram behind the blue desirability function shows the distribution of pK_i values for the compounds in the data set described in the text.

Related to this is the fact that experimental data on project compounds are often incomplete, with missing data points for compounds that have not yet been measured in one or more assays. This can be thought of as an extreme form of uncertainty, where the values of the corresponding properties are simply unknown within some reasonable bounds (although some level of expectation can be set from the prior distributions of similar compounds' measured properties [12]). It is not immediately clear how these missing data should be considered in MPO; one approach might be to set the desirability of missing data to 1 to avoid rejecting potentially good compounds, in the hope that subsequent measurements will reveal desirable properties. But, this may artificially promote compounds with many missing data points over those that have been progressed further, potentially wasting time and effort. In contrast, setting a lower desirability for missing data could artificially penalise compounds over those that have been measured, leading to missed opportunities, as discussed above. Most importantly, we would like to know when it would be valuable to measure missing data for a compound and 'fill in the blanks'.

To address these issues, the Probabilistic Scoring method [13] explicitly considers the uncertainty in each property value to calculate a score that represents the chance of success of a compound against a profile of property criteria, i.e. the probability of achieving ideal values for all properties. Each of the criteria in the profile is represented by a desirability function, as illustrated in Figure 2. In this case, the desirability of a property value is considered in terms of the risk associated with failure to meet the ideal outcome, i.e. the chance of success even if the compound's property does not meet the ideal criterion. Therefore, arbitrary weights for each property are not required in the calculation of the score and the importance of each criterion is defined as the difference between the ideal desirability of 1 and the lowest possible desirability for that property. A desirability function for which the lowest desirability is close to or equal to 0 represents a critical property criterion, i.e. the chance of success if the property value is poor will be very low. However, if the lowest desirability for a property is close to 1 (the ideal), this means that the criterion is 'nice to have' but failure to meet the criterion would not, by itself, have a significant impact on a compound's chance of success.

In keeping with the mathematics of probabilities, the overall score in Probabilistic Scoring is calculated using a multiplicative approach. However, to avoid giving too much weight to uncertain data, the uncertainty in each value is explicitly taken into account to consider the *probability* of achieving the ideal outcome for each property, as illustrated in Figure 3. In addition to the score for each compound, the uncertainty in the score is also calculated, as shown by the error bars in Figure 4. This makes it immediately clear when compounds can be distinguished by the data or where more precise data or an additional criterion is required to make a confident decision.

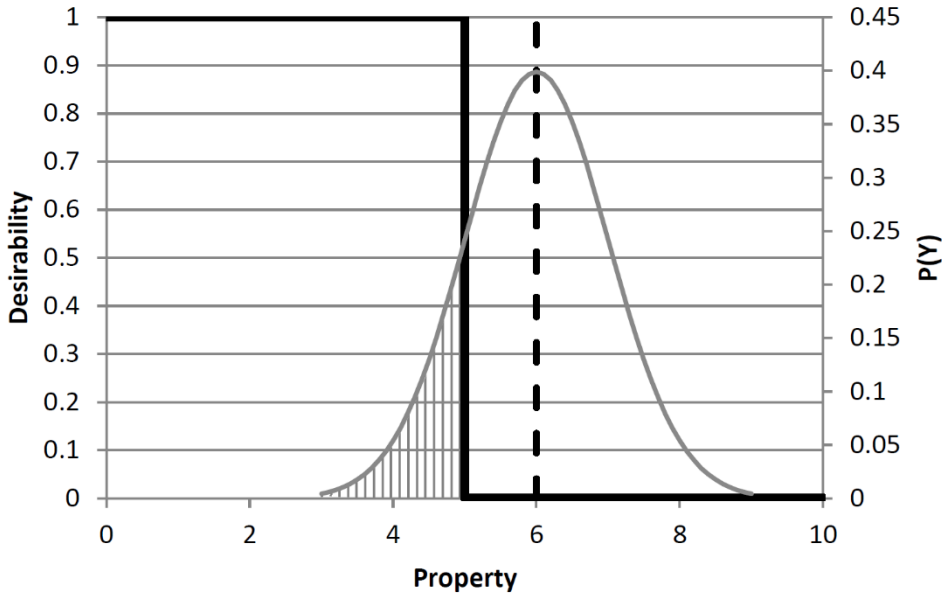


Figure 3 Illustration of the importance of uncertainty for a simple threshold desirability function (bold line), corresponding to a simple filter with a criterion of <4.0. The value of 6.0, determined for a compound's property, is indicated by the vertical dashed line and the uncertainty in this value by the grey bell curve (a Gaussian distribution with a standard deviation of 1.0). This indicates that, although the property value appears to be undesirable, there is a significant probability, shown by the hatched region, that it will meet the criterion. Therefore, it would be inappropriate to give the compound a desirability of 0 for this property and, in this case, the desirability would be 0.16, which is the probability that the compound will meet the criterion for this property.

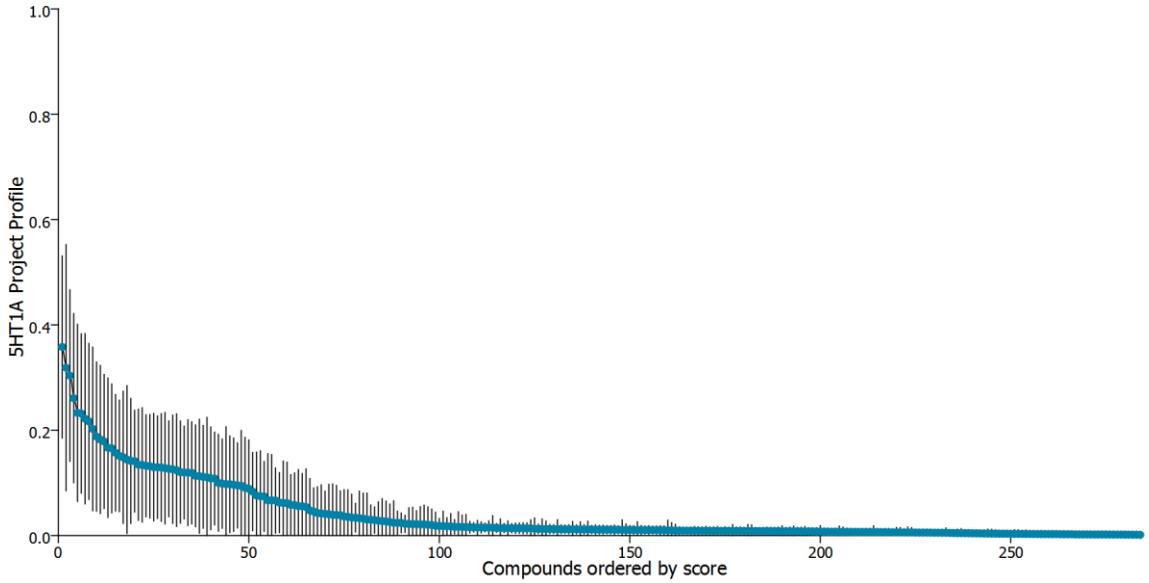


Figure 4. The results of probabilistic scoring for the data set described in the text, using the profile of property criteria shown in Figure 2. The compounds are ordered from left to right along the x-axis in order of their score and the overall score for each compound is plotted on the y-axis. The 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 clearly seen that the highest-scoring compound cannot be confidently distinguished from, approximately, the top 40 compounds in the data set.

This is particularly valuable in the case of missing experimental data, where the impact of these unknown values can be clearly seen. For example, if a compound had ideal outcomes with high confidence for all properties except for one critical property for which the data point was missing, this would be indicated by a moderate score; such a compound would be given lower priority than a compound that was known to be ideal for all properties but higher than a compound that was known to 'fail' for one or more properties. However, the score will have large error bars because, if the value of the missing data point was found to be good, the compound would be 'perfect' but, if the value was poor for this critical property, the overall chance of success would be low. This clearly highlights that it would be valuable to measure this missing data point. On the other hand, if a compound had poor outcomes for several properties and was missing data for a property, the score would be low with small error bars, because even a good value for this property would not 'rescue' such a compound; the effort of measuring this data point would likely be wasted.

MPO is applied throughout the drug discovery process, for example: in triage of early hits, to identify those most likely to provide access to series with good ADME and physicochemical properties; in hit-to-lead to quickly target high quality lead series with the best chance of success in lead optimisation; guiding *de novo* design to explore novel optimisation strategies; considering scaffold hopping strategies to overcome issues in lead optimisation; and in the prioritisation of lead compounds for detailed *in vivo* studies [3] [10] [14].

In this paper, we will use the Probabilistic Scoring method to illustrate the application of sensitivity analysis, but the principles apply equally to other MPO methods.

Sensitivity Analysis

Sensitivity analysis has been applied in other areas of modelling and simulation in computational chemistry and chemoinformatics. For example, the application of physiologically based pharmacokinetic (PBPK) simulations can be used to predict a compound's *in vivo* pharmacokinetic (PK) parameters based on *in vitro* measurements of properties such as metabolic stability, membrane permeability and lipophilicity. Analysis of the sensitivity of the predicted PK parameters to these properties can identify directions for compound optimisation; e.g. to improve *in vivo* half-life, is it more important to improve metabolic stability to reduce clearance or increase lipophilicity and hence volume of distribution [15]? Sensitivity analysis may also be applied in QSAR modelling to identify the most influential descriptors to include in a model [16].

To assess the sensitivity of the choice of compounds from a data set to the multi-parameter property profile used to prioritise the compounds, we consider the way that the ranking of the compounds varies when we make small changes to the property criteria and their importances [17]. When analysing the impact of changes in the value of a property criterion, we consider shifts in the corresponding desirability function, as illustrated in Figure 5. We use an equivalent approach to also consider the impact of changes to the importance of the criterion.

To calculate the sensitivity score due to a perturbation in a chosen property criterion, we start by computing the list of scores, $L_{original} = \{s_1, s_2, \dots, s_n\}$, of the *n* compounds in the data set, using the original scoring profile. Next, leaving all other property criteria as they were, we perturb the desirability function for the chosen property and then compute a new list of scores $L_{new} = \{t_1, t_2, \dots, t_n\}$ of the compounds in the data set. These scores may be calculated with any MPO method based on desirability functions and in the examples in this paper we have used the Probabilistic Scoring method described above.

We then compute Spearman's rank correlation coefficient between $L_{original}$ and L_{new}:

$$\boldsymbol{\rho} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}},$$

where $\{x_1, \dots, x_n\}$ is the list of ranks for each compound in $L_{original}$ and $\{y_1, \dots, y_n\}$ is the list of ranks for each compound in L_{new}.

However, if the order of two compounds only changes within the uncertainties in the scores, this is not significant; the change is 'within the noise'. In other words, the perturbation to the scoring profile doesn't change the conclusion that the compounds cannot be confidently distinguished (see Figure 6 for an illustration of this effect). Therefore, to account for the effect of uncertainty in the scores, we adjust the standard computation of Spearman's rank correlation coefficient by assigning the maximum possible correlation contribution to compounds with old and new scores that are not statistically significantly different when their

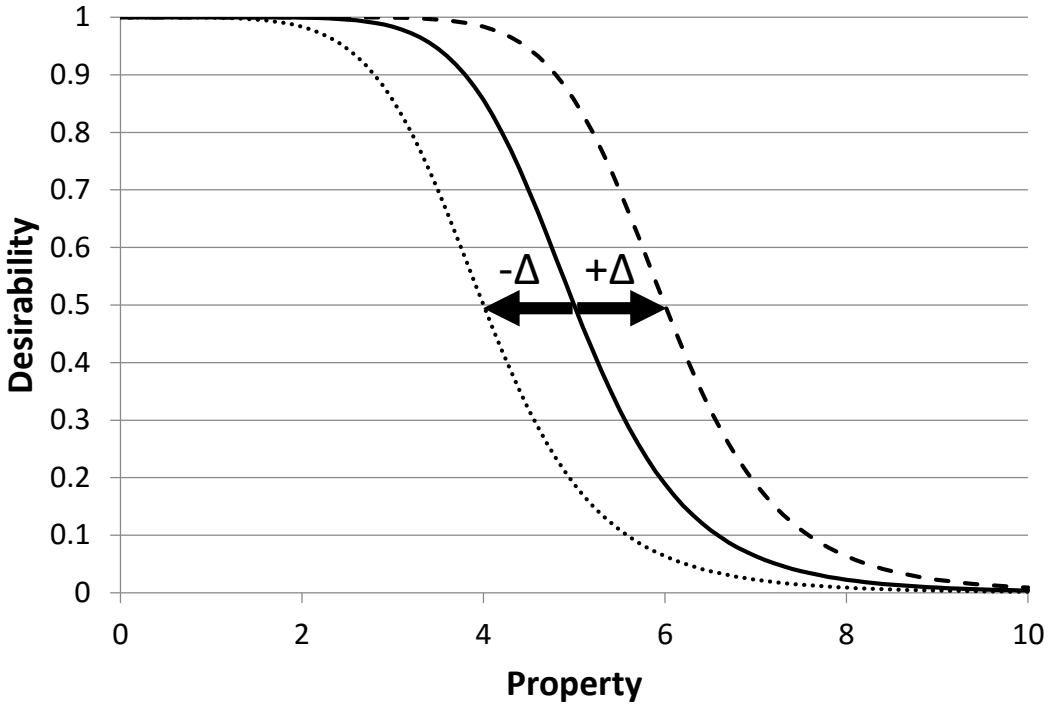


Figure 5. Illustration of shifts in a sigmoidal desirability function of Δ in positive and negative directions. The solid black line shows the original desirability function, the dashed line the same desirability function shifted by Δ in the positive direction and the dotted line the desirability function shifted by Δ in the negative direction.

uncertainties are taken into account. Specifically, if compound *i* has the original score s_i and new score t_i, we calculate $P(|s_i - t_i| > 0)$, assuming the individual scores are normally distributed. If this probability is below a specified significance threshold (for example 0.75 [18]), we consider the score change to be insignificant. In this case, if the compound has original rank x_i, we change its new rank from y_i to $x_i - (\bar{x} - \bar{y})$, i.e. the compound is given the same rank translated by the difference between the means of the original and new sets of ranks.

Furthermore, in practice, we are only interested in whether the order of the top-ranked compounds (i.e. those compounds with the highest *original* scores) changes; it is not relevant if the order of compounds which would have been rejected changes. Therefore, only the top-scoring compounds are considered in the computation of the correlation coefficient (for example, the top 10% or another number relevant to the project). The sensitivity score for the single shift to the scoring function or change in the importance is then defined to be $1 - \rho$. While it is theoretically possible for the correlation coefficient to be negative, in practice this is very unlikely for the perturbations made in this analysis. Therefore, a sensitivity value between 0 (no sensitivity) to 1 (highly sensitive) is reported for ease of interpretation.

To calculate the overall sensitivity with respect to the value of a property criterion, we repeat this calculation for multiple, positive and negative, shifts in the desirability function within a 'window'. The window of shifts is symmetric around 0 and the size of the window is defined by a fraction of the total range of the property in the data set (for example 50%). The overall sensitivity score for the value of the property criterion is then defined to be the maximum sensitivity score for shifts within this window. Similarly, the overall sensitivity for the importance of a criterion is defined as the maximum sensitivity for importance values within a window from -0.25 below, and 0.25 above, the original importance (subject to a minimum importance value of 0 and maximum of 1).

The results are therefore overall sensitivities between 0 (no sensitivity) and 1 (highly sensitive) for the value and importance of each property criterion in the profile.

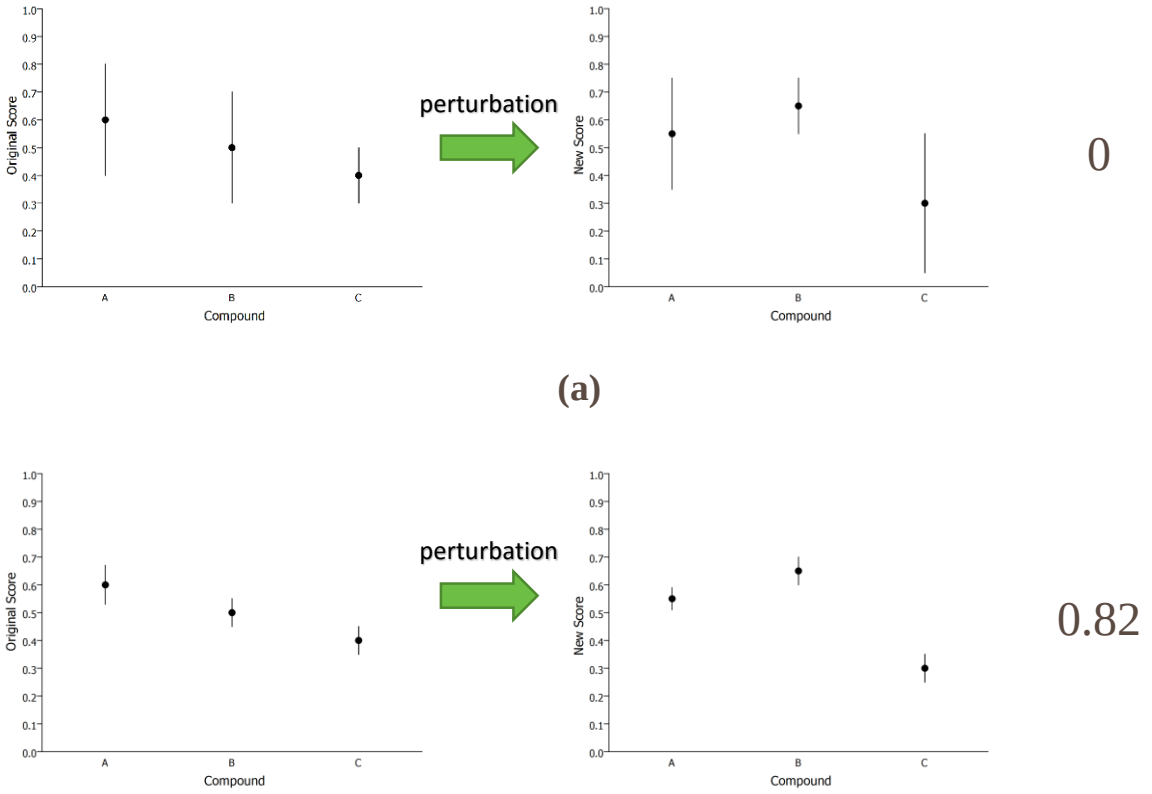


Figure 6. Illustration of the effect of uncertainty on the sensitivity due to a perturbation. The plots on the left show the scores of three compounds A, B and C calculated with the original scoring profile. The plots on the right show new scores calculated with a new scoring profile created by applying a perturbation to the original scoring profile. In each plot the error bars show the standard deviations of the scores. In (a) the error bars overlap significantly, indicating that the compounds cannot be confidently distinguished, therefore the sensitivity due to this perturbation is small, i.e. the change in rank order is within the 'noise' in the data. However, in (b) the error bars are smaller and hence this reordering is statistically significant and therefore the rank order of the compounds is sensitive to the perturbation.

Discussion and Conclusion

Visualisation and Interpretation

There are three questions we would like to address using sensitivity analysis:

- To which criteria values or their importances in the profile (if any) is the selection of compounds sensitive?
- If we are uncertain about the 'right' value for a criterion or its importance, is the selection sensitive within a 'reasonable' range of possible values?
- If the selection of compounds is sensitive to a change in the profile, which compounds are affected by the change?

The first of these questions can be answered by considering the overall sensitivity values for the criteria values and their importances. A sensitivity value near to one indicates a highly sensitive parameter, i.e. the choice of compounds will change significantly for small changes to that parameter. In this case the corresponding criteria value or its importance should be considered carefully to ensure that it is appropriate. If there is uncertainty about the most appropriate value for the parameter, alternative compounds that would be selected by different profiles with 'reasonable' values of that parameter should be considered. Conversely, if the selection of

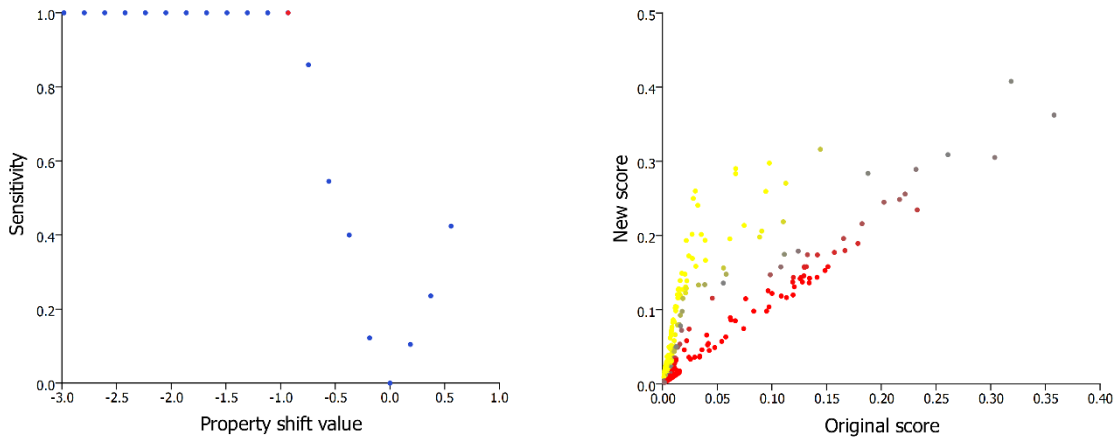
compounds is not sensitive to any of the parameters of the scoring profile, small changes to the profile would not affect the selection of compounds, hence the project can proceed with confidence.

It can also be instructive to visualise a graph of the sensitivity against perturbations in a property criterion or its importance, as shown in Figure 7a and Figure 8a. These graphs indicate the range of a parameter over which the choice of compound is insensitive. A highly sensitive parameter will appear as a very steep curve around the unperturbed value, as illustrated in Figure 7a. On the other hand, if the plot is relatively shallow (such as the example in a) and all reasonable values of the parameter lie within the range with low sensitivity, the selection of compounds can also proceed with confidence.

Finally, for a sensitive parameter, the impact of a perturbation on the priority of individual compounds can be easily assessed by plotting the new compound scores for the perturbed profile against the original scores. Those points which deviate significantly from the identity line are those for which the scores have changed significantly due to the perturbation. Again, particular attention should be paid to changes to the top-ranked compounds, because these are the changes that would affect any selections made. An example of this is shown in Figure 7b.

In a similar manner to MPO itself, sensitivity analysis may be applied throughout the drug discovery process. It is perhaps most valuable to analyse the sensitivity of decisions when there is the greatest flexibility to choose between different chemical series, for example when selecting hit or lead series, because making the most appropriate choice at this stage will have the greatest impact on downstream progress. This is illustrated in the example below. Once locked into a potent series, it can become more difficult to make significant changes to synthetic strategy, but even late in a project sensitivity analysis can reveal when the selection criteria being applied may be distorting candidate selection decisions or identify potential back-up series.

Finally, it is important to note that the sensitivity of the criteria and their importances in a property profile will depend on the data set to which it is applied. Therefore, as new compounds are considered and new data are generated, it is useful to repeat the sensitivity analysis on a regular basis to check if new avenues for exploration are revealed. Fortunately, the computational tools to perform this analysis are quick and easy to apply and interpret.



(a) **(b)**

Figure 7. Graphs illustrating the sensitivity of the rank order of compounds in the data set described in the text to perturbations in the value of the pK_i criterion in the scoring profile shown in Figure 2. (a) shows the sensitivity plotted against the shift in the value of the criterion, both positive and negative. The high curvature around a shift of zero shows that the rank order of the highest scoring compounds is very sensitive to changes in the value of this criterion. (b) shows the change in the scores of the compounds in the data set due to a negative shift of one log unit (highlighted by the red point in (a)) by plotting the compound scores calculated with the original profile against new scores calculated with the profile in which the pK_i criterion has been perturbed. The colours indicate the significance of the change in rank order; yellow represents a significant positive change; grey, no significant change and red a significant negative change. Here we can see that the top-ranked compounds are significantly changed by this perturbation.

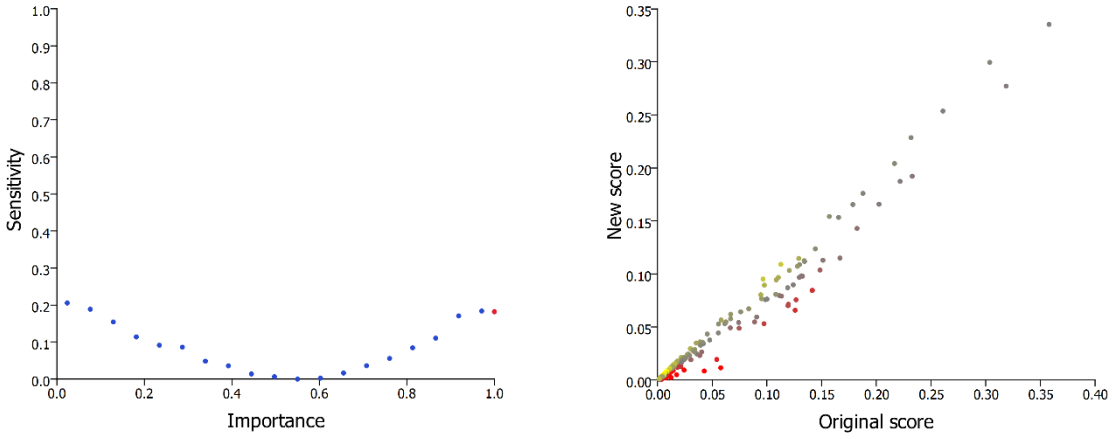


Figure 8. Graphs illustrating the sensitivity of the rank order of compounds in the data set described in the text to perturbations in the importance of the BBB log([brain]:[blood]) criterion in the scoring profile shown in Figure 2. (a) shows the sensitivity plotted against different levels of importance assigned to the criterion (the original value is 0.55). The low curvature around the original value shows that the rank order of the highest scoring compounds is not sensitive to changes in the importance of this criterion. (b) shows the change in the scores of the compounds in the data set by plotting the compound scores calculated with the original profile against those calculated with the profile in which the BBB log([brain]:[blood]) criterion has been increased to 1 (highlighted by the red point in (a)). The colours indicate the significance of the change in rank order; yellow represents a significant positive change; grey, no significant change and red a significant negative change. Here we can see that the top-ranked compounds are not significantly changed by this perturbation.

Example Application

As an example, the methods described above have been applied to a data set of 284 compounds for which the inhibition of the intended therapeutic target (pK_i) has been measured and predictions have been made using QSAR models for solubility (logS), human intestinal absorption (HIA category based on a cut-off of 30%), logarithm of the partition coefficient between octanol and water (logP), blood-brain barrier penetration (BBB log([brain]:[blood])), P-glycoprotein transport (P-gp category), inhibition of the hERG ion channel (hERG pIC_{50}), affinity for the CYP2C9 and CYP2D6 isoforms of cytochrome P450 (2C9 pK_i and 2D6 category respectively) and plasma-protein binding (PPB category based on a cut-off of 90%). Details of the experimental data and predictive models used in this example are given in the Experimental section and the data set is provided as Supporting Information.

These compounds were scored using the profile shown in Figure 2, which was determined based on the subjective opinions of the project team. The objective was to identify a high quality lead compound for a CNS target and oral administration. Therefore, suitable potency, solubility, absorption, lipophilicity and blood-brain barrier penetration where given the highest priorities. Avoiding active transport by P-glycoprotein was also desirable because efflux across the blood-brain barrier can limit brain exposure. Furthermore, it was considered to be desirable to avoid inhibition of hERG and major Cytochrome P450 isoforms to reduce the risk of cardiotoxicity and drug-drug interactions. Low plasma-protein binding is also ideal in order to maximise the free drug concentration in circulation, but this was given the lowest importance because the 90% classification boundary of this model is not indicative of a critical issue. The Probabilistic Scoring method was applied to the data and associated uncertainties, resulting in the distribution of scores shown in Figure 4. All of the QSAR predictions, Probabilistic Scoring and sensitivity analysis calculations and visualisations were generated with the StarDrop software platform [19].

Given the subjective nature of the property profile, it is important to consider the possible impact of the chosen criteria on the selection of compounds and hence the direction of the project. Table 1 shows the sensitivity scores for each of the criteria values and their importances. From this, it is immediately apparent that the

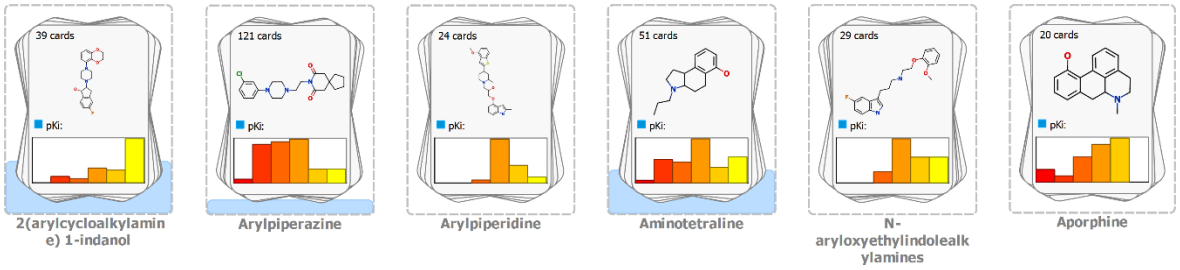
selection of compounds is highly sensitive to the value chosen for the pK_i criterion (the desirability function shown in Figure 2). A plot of the sensitivity of compound selection to perturbations in the value of this criterion is shown in Figure 7a. From this, we can see that a reduction of this criterion by one log unit would dramatically change the order in which compounds would be chosen, as shown in Figure 7b. In other words, compromising on the range of acceptable potency would give access to compounds with significantly better predicted ADME properties.

Table 1 Sensitivities of the value and importance for each of the property criteria shown in the profile in Figure 2, calculated for the data set of 284 compounds described in the text using the method for sensitivity analysis described in the Methods section. Note that no sensitivity can be calculated for the value of a categorical property because the boundary of the categorisation is fixed.

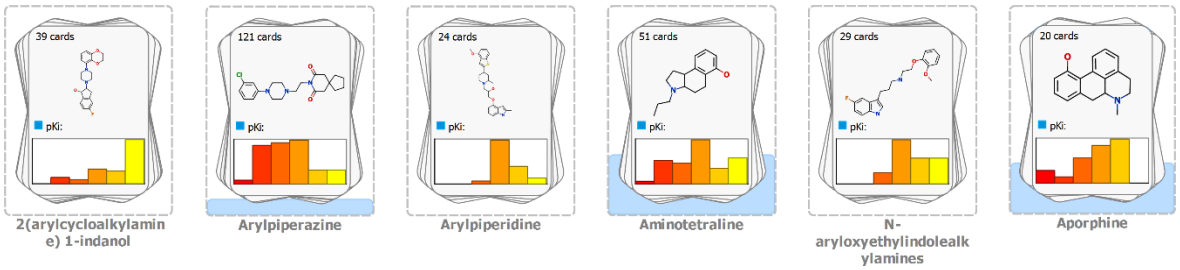
Property	Value Sensitivity	Importance Sensitivity
pKi	1.00	0.25
logP	0.33	0.21
logS	0.26	0.00
P-gp category	N/A	0.17
BBB log([brain]:[blood])	0.16	0.08
hERG pIC_{50}	0.10	0.07
BBB category	N/A	0.09
PPB90 category	N/A	0.02
2D6 affinity category	N/A	0.01
2C9 pK_i	0.01	0.00
HIA category	N/A	0.00

Furthermore, if the objective were to select 30 compounds (approximately 10% of the original data set), Figure 9 shows that this compromise would change the chemical series selected for progression. Compromising slightly on potency, at this stage of the project, would result in the exploration of a different chemical series which is more likely to achieve better ADME properties. In this case, it may be appropriate to select some examples of the alternative chemical series to consider a broad range of options and validate the best strategy for progression.

On the other hand, BBB log(brain]:[blood]) is an example of a parameter to which the choice of compounds is insensitive. This is shown in Figure 8a, where perturbations in the importance of the BBB log([brain]:[blood] criterion from an original value of 0.55 produce a very shallow curve. Even if the importance of this property criterion were changed to 1 (critical), this would have very little impact on the order of compounds, as shown in Figure 8b. Therefore, there is no need to be concerned over the choice of the most appropriate importance for this criterion. In other words, the top-scoring compounds in this data set are not distinguished by their predicted blood-brain barrier penetration; this may, of course, be different if the scoring profile was applied to another data set.



(a)



(b)

Figure 9. Illustration of the change in the selection of the 30 top-ranked compounds from the data set described in the text, due to a perturbation of the scoring profile shown in Figure 2. The chemical series in the data set are shown as 'stacks' of 'cards', on top of which are shown the number of compounds in each series, a representative structure and a histogram showing the distribution of the pK_i values for the compounds in the series. The proportion of compounds selected in each series is shown by the filled blue region surrounding the stack (a) shows distribution of the 30 top-ranked compounds when scored with the original scoring profile. (b) shows the distribution of the 30 top-ranked compounds when scored with the perturbed profile in which the value of the pK_i criterion has been shifted negatively by one log unit. Here we can clearly see that the change in the pK_i criterion results in a significant change in the chemical series that would be progressed.

Conclusion

In this paper we have described a method to rigorously assess the sensitivity of the decisions made regarding the progression of compounds in the context of a project's specific compound property requirements. It is important to ensure that the decisions made are robust to the, sometimes subjective, choices for property criteria used for the selection of compounds.

We have also illustrated that, when prioritising compounds and assessing the stability of the results, it is important to consider the impact of uncertainty. All data in drug discovery, whether experimentally measured or predicted have significant uncertainty due to experimental variability and statistical error. When data do not support the ability to distinguish between compounds, strategies should be considered that explore a range of options to ensure that valuable opportunities are not rejected, with a view towards generating more confident data where necessary.

If the direction of a project is found to be sensitive to the value or importance of a property criterion, what should be done? Consider the value or importance of the property criterion carefully and, if you are confident that this has been set correctly, you can proceed. However, if you are uncertain about what the 'right' value of the parameter should be, a more rigorous strategy may be to progress some of the compounds selected by different values, for more detailed experimental investigation, to determine the most appropriate value to use in the future.

Experimental

In order to avoid issues with confidentiality of structures and data, the data set used in the illustrative example above was derived from the public domain ChEMBL database [20] and represents compounds for which inhibition (pK_i) had been measured for the 5-hydroxytryptamine receptor type 1A (5-HT_{1A}). The activities span 5 orders of magnitude with a mean pK_i value of 7.6 and a standard deviation of 1.0 log units, as illustrated in Figure 10. The uncertainty in the experimental values was estimated to be 0.3 log units. The compounds in the data set represent 6 chemical classes as illustrated in Figure 9.

The remaining compound properties were calculated with QSAR models in the StarDrop software platform [19]. The models were built using a variety of QSAR modelling methods and descriptors including whole molecule properties such as logP, polar surface area, MW, charge and flexibility and counts of structural fragment descriptors. Details of the models, including the method used, numbers of compounds in the training and test sets and external validation results are shown in Table 2.

Table 2 Summary of the QSAR models used in the illustrative example. For each model, the method used, numbers of compounds in the training and test set ($N_{training}$ and N_{test}) and the validation results for the independent test are provided. For regression models, the coefficient of determination (R^2) and root-mean-square error (RMSE) are shown; for classification models the classification accuracy is given for each class.

Model	Method	$N_{training}$	N_{test}	Independent Test	
Regression models				**R^2**	**RMSE**
logP	Radial basis function	6887	2950	0.92	0.44
logS	Radial basis function	2650	663	0.82	0.80
BBB log([brain]:[blood])	Radial basis function with genetic algorithm feature selection	359	75	0.72	0.36
hERG pIC_{50}	Gaussian Process	135	33	0.72	0.64
2C9 pK_i	Random forest regression	105	25	0.64	0.60
Classification Models				**Accuracy**	
P-pg category	Random forest	205	51	79% (yes), 82% (no)	
BBB category	Random forest	101	100	83% (+), 91% (-)	
PPB90 category	Random forest	775	332	81% (high), 87% (low)	
2D6 affinity category	Decision tree (4 class)	168	45	63% (low), 50% (medium), 59% (high), 50% (very high)	
HIA category	Decision tree	252	245	99% (+), 66% (-)	

Each QSAR model is associated with a domain of applicability, defined by the diversity of the training set used to build the model. The position of a predicted compound in relation to the domain of applicability is reflected in the uncertainty in the predicted value. In some cases, this is derived from the root-mean-square or classification errors of compounds in the independent validation set of the model both inside and near to the domain of applicability, as defined by a Hotelling's T^2 test in the descriptor space of the model. For compounds lying far from the domain of applicability, a very high uncertainty is reported to reflect the lack of confidence in such a prediction. For modelling methods, such as random forests [21] or Gaussian Processes [22], the uncertainty may be estimated on an individual compound basis. These uncertainties are naturally taken into consideration by the Probabilistic Scoring and sensitivity methods, as described above.

Acknowledgements

The authors would like to thank Dr Peter Hunt for reviewing the draft manuscript and for his helpful suggestions.

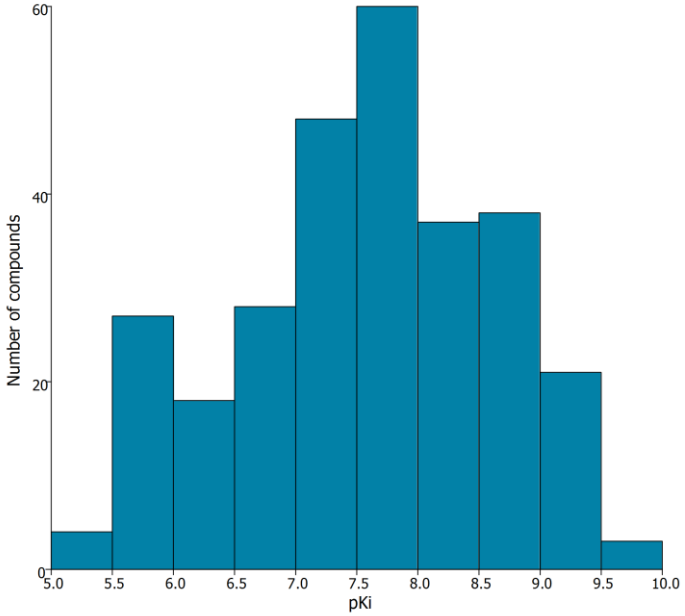


Figure 10 Histogram showing the distribution of measured inhibition values (pK_i) for the data set used in the illustrative example.

References

1 Yusof I, Shah F, Hashimoto T, Segall MD, Greene N. Finding the rules for successful drug optimization. Drug Discovery Today. 2014;19(5):680-687.

2 Petit J, Meurice N, Kaiser C, Maggiora G. Softening the Rule of Five--where to draw the line? Bioorg. Med. Chem. 2012;20(18):5343-51.

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

4 Nicolaou CA, Brown N. Multi-objective optimization methods in drug design. Drug Discovery Today: Technol. 2013;10(3):e427-e435.

5 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.

6 Segall MD, Champness EC. The challenges of making decisions using uncertain data. J. Comput.-Aided Mol. Des. 2015;29(9):809-816.

7 Harrington EC. The desirability function. Ind. Qual. Control. 1965;21(10):494-498.

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

9 Wager TT, Hou X, Verhoest PR, 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.

10 Segall MD. Advances in multiparameter optimization methods for de novo drug design. Expert Opin. Drug Discovery. 2014;9(7):803-817.

11 Debe DA, Mamidipaka RB, Gregg RJ, Metz JT, Gupta RR, Muchmore SW. ALOHA: a novel probability fusion approach for scoring multi-parameter drug-likeness during the lead optimization stage of drug discovery. J. Comput.-Aided Mol. Des. 2013;27(9):771-82.

12 Segall MD, Chadwick A. Making priors a priority. J. Comp.-Aided Mol. Des. 2010;24(12):957-960.

13 Segall M, Beresford A, Gola J, Hawksley D, Tarbit M. Focus on success: Using in silico optimisation to achieve an optimal balance of properties. Expert Opin. Drug Metab. Toxicol. 2006;2(2):325-337.

14 Green DVS, Segall M. Chemoinformatics in lead optimization. In: Bajorath J, editor. Chemoinformatics in drug discovery. Hoboken, N.J.: John Wiley & Sons, Inc.; 2013. p. 149-178.

15 Peters SA. Variability, uncertainty and sensitivity analysis. In: Peters SA, editor. Physiologically-based pharmacokinetic (PBPK) modeling and simulations: Principles, methods, and applications in the pharmaceutical industry. Hoboken, NJ: John Wiley & Sons, Inc.; 2012. p. 161-181.

16 Guhama R, Jurs PC. Interpreting computational neural network QSAR models: A measure of descriptor importance. J. Chem. Inf. Model. 2005;43(3):800-806.

17 Segall MD, Yusof I, inventors. Sensitivity analysis tool for multi-parameter selection. 2015 Dec 29. US patent 9,224,098.

18 This value is provided as an example and may be modified if required. The suggested value of 0.75 was determined empirically based on typical drug discovery data to achieve a good dynamic range for the calculated sensitivity values. A high value of this probability means that many compounds will be considered statistically indistinguishable, unless high precision data is available, hence sensitivity values will typically be low because changes in the selection order will be lost in the 'noise'. Conversely, a low value of this probability means that scores will be considered different even if there is a significant overlap between their associated error bars; hence the sensitivity values will typically be high and not reflect the lack of statistical significance in the reordering of compounds.

19 Optibrium. [Internet]. [cited 2016 February 5]. Available from: http://www.optibrium.com/stardrop.

20 ChEMBL. [Internet]. [cited 2016 February 5]. Available from: https://www.ebi.ac.uk/chembldb/.

21 Svetink V, Liaw A, Tong C, Culberson J, Sheridan R, Feutson B. Random forest: A classification and regression tool for compound Ccassification and QSAR modeling. J. Chem. Inf. Comput. Sci. 2003;43:1947-58.

22 Obrezanova O, Csanyi G, Gola J, Segall M. Gaussian processes: a method for automatic QSAR modeling of ADME properties. J. Chem. Inf. Model. 2007;47:1847-1857.