

Multi-Parameter Optimization: Identifying high quality compounds with a balance of properties

Matthew D Segall

Optibrium Ltd., 7226 Cambridge Research Park, Beach Drive, Cambridge, CB25 9TL, UK

Email: matt.segall@optibrium.com

Keywords: "Multi-parameter optimization", "Multi-objective optimization", "Multi-dimensional optimization", "ADMET", "Lead Identification", "Lead Optimization", "Probabilistic scoring", "Pareto optimization", "Desirability function"

Abstract

A successful, efficacious and safe drug must have a balance of properties, including potency against its intended target, appropriate absorption, distribution, metabolism, and elimination (ADME) properties and an acceptable safety profile. Achieving this balance of, often conflicting, requirements is a major challenge in drug discovery. Approaches to simultaneously optimizing many factors in a design are broadly described under the term 'multi-parameter optimization' (MPO). In this review, we will describe how MPO can be applied to efficiently design and select high quality compounds and describe the range of methods that have been employed in drug discovery, including; simple 'rules of thumb' such as Lipinski's rule; desirability functions; Pareto optimization; and probabilistic approaches that take into consideration the uncertainty in all drug discovery data due to predictive error and experimental variability. We will explore how these methods have been applied to predicted and experimental data to reduce attrition and improve the productivity of the drug discovery process.

Preprint of article to appear in a special issue of Current Pharmaceutical Design.

Introduction

A successful drug that passes the hurdles of clinical trials to gain approval and a strong market position must exhibit a delicate balance of biological and physicochemical properties. Such a compound must, of course, be potent against its intended physiological target(s); however, it must also have appropriate pharmacokinetics to reach the site of the target at a sufficiently high concentration and for an appropriate duration via the intended route of administration. Furthermore, for the compound to be safely administered, it must avoid unintended side-effects, drug-drug interactions and non-specific or idiosyncratic toxicities at the therapeutic dose. The goal of drug discovery is to identify a successful compound as efficiently as possible. But, as the history of drug discovery has proved, this is a challenge of significant proportions [1].


This task is made even more difficult by the fact that, in drug discovery, data on the behavior of the compound in the ultimate target patient population, i.e. humans, is not available. This has led to the development of a plethora of *in silico*, *in vitro*, and *in vivo* animal models from which we can (hopefully) infer the likely *in vivo* efficacy, disposition and safety of a compound in humans. These include models for the prediction and measurement of potency and selectivity against molecular targets or off-targets; absorption, distribution, metabolism and elimination (ADME) properties; cell-based measurements of pharmacological activity and toxicity; and animal models of pharmacology, pharmacokinetics and toxicity. The cost and throughput of these techniques vary, from *in silico* methods which typically have the lowest cost and highest throughput, through *in vitro* and cell-based assays to lengthy and expensive *in vivo* studies, the use of which we would also like to minimize for ethical reasons. Therefore, drug discovery is a process of simultaneously optimizing all of these factors as compounds are designed, synthesized and progressed through a cascade of assays to accumulate data.

This balancing act is difficult to achieve through a purely intellectual process. Psychologists have repeatedly demonstrated that people are very poor at making decisions based on complex and uncertain data when there is a lot at stake, such as in drug discovery. Several biases in decision-making (described as cognitive biases) have been identified that can detrimentally affect efficiency and productivity in drug discovery. A detailed discussion of some of these, with examples, may be found in [2]; however two illustrative examples are:

- Confirmation bias: The tendency to seek data that confirms a pre-formed hypothesis, rather than perform experiments designed to yield results to challenge the hypothesis. This can lead to a premature focus on a small range of options, which may lead to missed opportunities or late stage failures of compounds that have been progressed too far in the search for the one piece of data that would prove the point.
- Excess focus on certainty: The tendency to seek additional data to be ‘absolutely certain’ of a critical factor, even when this data adds little value at a high cost. Often a more significant increase in the confidence around a slightly less important factor may have a greater effect on the overall chance of success. This can lead to inefficient use of resources when considering multiple property requirements and to late stage, expensive failures.

The historical evidence regarding the attrition and productivity of pharmaceutical research and development supports this observation. The increasing complexity and volume of data being generated in drug discovery has not improved success rates in development – 11% in 2000 [3] versus 12% in 2010 [1] – while the cost per marketed drug has continued to escalate – from an estimated fully capitalized cost of \$802M in 2001 [4] to \$1,778M in 2010 [1] – and productivity, as measured by the number of registered new chemical entities, has fallen [5]. There are a wide range of theories regarding the underlying cause of these effects, but it is safe to conclude that generating additional, early-stage data has not resulted in the improvements anticipated in the outcomes.

Fortunately, we may learn from other fields that face the same need to balance many factors in the design of a successful solution. These fields range from engineering disciplines, such as aerospace or automotive design, to economics. The resulting methods are commonly described under the broad term “Multi-parameter Optimization” (MPO) or sometimes also “Multi-dimensional Optimization” (MDO) or “Multi-objective Optimization” (MOOP). For convenience, we will use the term MPO to describe all of the methods in this review.



There is a significant difference between applications of MPO methods to drug discovery and other fields, in particular engineering. This relates to the quality of the data available on the potential designs or prototypes from which a selection must be made. In an engineering discipline, characteristics may commonly be measured to accuracies within parts per million or predicted computationally to within a fraction of a percent. This may be contrasted with drug discovery where measured properties, such as IC_{50} or K_i values, may have an experimental variability of a factor of two, while predictions may have statistical uncertainties of an order of magnitude. This dramatically increases the challenge because, even if an ideal compound exists among the available options, we cannot expect to identify it with absolute confidence, thus running the risk of missing opportunities for high quality drugs [6].

In our research into the requirements for an ideal MPO method for drug discovery, we identified the following factors that should be taken into account:

- **Interpretability:** The property criteria and their impact on compound priority should be easy to understand. A 'black box' method that does not provide an easy way to understand why a compound has been classified in a given way is likely to be discounted. Furthermore, a 'black box' does not provide any guidance on the way one should go about making improvements in order to increase the chance of success.
- **Flexibility:** Each project will have a different set of property criteria depending on the therapeutic objectives of the project, intended route of administration and competitive conditions in the market. The project team should be able to define appropriate criteria based on their experience or historical evidence.
- **Weighting:** The project team should be able to assign different weights to each property criterion, as different criteria will have different degrees of importance to the outcome of the project. For example identifying a compound that is potent against the intended target is critical, while other properties will be less important, particularly early in a project when there is an opportunity for redesign to overcome liabilities.
- **Uncertainty:** It is important to avoid rejecting potentially valuable compounds based on a property value that fails to meet a criterion if that value has a high level of uncertainty. The opportunity cost of incorrectly rejecting a good compound may be very high, particularly when the range of alternative options is limited.

Coincidentally, it seems that the development of a suitable MPO approach for drug discovery is itself an MPO problem!

One common question is, "Can't this be easily solved by visualization of the data?" While visualization is necessary to understand and communicate results, it is not sufficient to allow conclusions to be easily drawn, given the complexity of the data at hand. One common approach is to plot multi-dimensional data, for example on a three dimensional graph with additional parameters shown by the colors and sizes of the points. An alternative is a 'traffic light' view where the properties of each compound are shown in a table and colored according to whether they 'pass' (green), 'fail' (red) or are 'close' (yellow) to the relevant criterion. However, even with only five-dimensional data, it is difficult to confidently draw a conclusion from these visualizations even before we consider the relative importance of each property or the uncertainty in the data. An MPO method helps a project team to define a set of criteria and use this pro-actively to guide their decisions to quickly target high quality compounds [7].

In this review, we will explore a range of different MPO approaches that have been applied to drug discovery and compare their strengths and weaknesses relative to the requirements described above. The methods that we will discuss in order of increasing sophistication, include 'rules-of-thumb' that provide chemists with guidelines for compound characteristics, simple pass/fail filters, Pareto optimization, desirability functions and probabilistic scoring, which brings together all of the requirements discussed above. We will also consider the role of chemical diversity to mitigate risk when selecting compounds for further investigation. Finally, we will illustrate some of the methods using examples taken from the literature before drawing our conclusions.

Rules of Thumb

Perhaps the most common approach used to consider the quality of compounds relative to criteria beyond potency are 'rules of thumb' that provide guidelines regarding desirable compound characteristics. The best known is undoubtedly Lipinski's Rule of Five (RoF) [8], which proposes criteria for four basic characteristics that Lipinski identified as being satisfied by the majority of orally absorbed compounds, namely:

- Molecular Weight (MW) < 500
- Logarithm of the octanol:water partition coefficient ($\log P$) < 5
- Number of Hydrogen Bond Donors (HBD) < 5
- Number of Hydrogen Bond Acceptors (HBA) < 10

Subsequently, several other rules have been proposed, for example Veber *et al.* [9] identified that most of the 1100 compounds they studied with oral bioavailability of greater than 20% in rats had less than 10 rotatable bonds and a Polar Surface Area (PSA) of less than 140 \AA^2 . However, Lu *et al.* [10] repeated this study with a set of 434 compounds and showed that the criteria depended on the method used for calculation, providing one illustration of the need for flexibility in the criteria depending on the source of data.

Johnson *et al.* identified rules based on MW and the logarithm of the octanol:buffer partition coefficient at pH7.4 ($\log D$) to achieve permeability and metabolic stability [11]. In this case, rather than expressing these rules as criteria for the individual characteristics, Johnson *et al.* identified correlations that led them to express the rules in terms of a 'golden triangle' that defines an optimal region in ($MW, \log D$) space in which a compound should lie (illustrated in Figure 1).

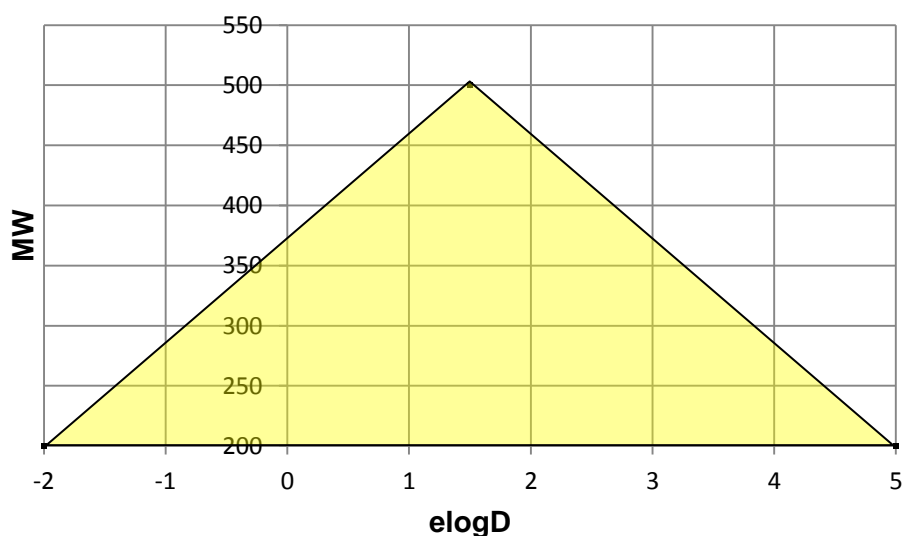


Figure 1. An illustration of the "Golden Triangle" [11] proposed by Johnson *et al.* Compounds within the shaded region in ($MW, \log D$) space were found to have a higher chance of achieving better outcomes for permeability and metabolic stability. This is a convenient visual rule-of-thumb for selecting compounds.

Other rules, involving parameters such as the fraction of carbons which are sp^3 hybridized [12] and the number of aromatic rings [13] have been proposed as measures of developability or likelihood of clinical success. Furthermore, Hughes *et al.* [14] studied the relationship between physicochemical properties and adverse events observed in *in vivo* toleration studies. They concluded that compounds with both calculated $\log P$ (clogP) > 3 and topological polar surface area (TPSA) < 75 \AA^2 had a significantly increased safety risk.

The undoubted popularity of these rules derives from their simplicity and interpretability, the first requirement for a good MPO method. It is very easy to calculate these characteristics and quickly check if a compound obeys these rules. Similarly it is easy to understand how to modify a compound that fails to meet these rules in order to improve its chance of success; it is clear how MW, HBD or HBA could be reduced and

chemists have a good understanding of the influence of chemical functionalities on lipophilicity. Therefore, these rules-of-thumb provide an easy approach to selecting compounds and guiding their redesign.

The main disadvantage of these rules-of-thumb is also due to their simplicity. There may be a tendency to over-interpret simple rules and apply them with too much rigor. For example, does a compound with a MW of 501 have a significantly worse chance of oral absorption than one with MW of 500? Indeed, Lipinski's original paper [8] suggested that two or more failures against the RoF criteria were required to significantly decrease the chance of oral absorption, so the rules were not intended to be applied individually.

These rules are derived from a review of historically successful drugs and are often treated as absolute rules that define 'drug likeness.' However, compounds for different therapeutic indications or routes of administration may require different characteristics or be more tolerant to violations of these rules. For example, there has been a tendency for the RoF to be considered as a definition of the conditions for 'drug likeness' when it is only based on analysis of the requirements for orally absorbed drugs. Drugs intended for topical, IV, inhaled or other routes of administration can violate some or all of the rules without a significant impact on their chance of success [15]. Therefore, the criteria and weight given to each of these rules-of-thumb should be defined or applied flexibly according to the therapeutic objectives of a project. Unfortunately, this is not a straightforward exercise, as careful statistical analysis of a large number of compounds is required to identify statistically significant criteria.

The majority of the characteristics used in these rules-of-thumb do not have any underlying uncertainty, as they are simple values calculated from the molecular structure. The principal exception to this is lipophilicity (logP or logD) which, if calculated, typically has an uncertainty (root-mean-square-error) of at least 0.5 log units. Therefore, care should be taken when drawing conclusions regarding compounds close to the criterion for lipophilicity.

Finally, we should consider the confidence in the 'prediction' by a rule-of-thumb. As an illustrative example, we applied the RoF to a set of 1191 marketed drugs labeled according to whether they have been approved for oral administration and the results are shown in Table 1. Although, one should be careful not to over interpret these results, we can see that passing the RoF is not a guarantee of finding an orally available compound. This is not surprising, as the RoF was derived from observations of absorption and other factors such as first pass metabolism can limit oral bioavailability. However, the specificity of the RoF is also low (21%), as more non-orally administered compounds pass the RoF than fail and a significant proportion of compounds that fail the RoF are orally administered.

Table 1. The results of applying Lipinski's Rule of Five to 1191 marketed drugs labeled as oral or non-oral according to their approved route of administration.

	RoF result	
	Pass (≤ 1 RoF Failure)	Fail (> 1 RoF Failure)
Oral	709	59
Non-oral	333	90

In summary, rules-of-thumb can provide very convenient and easily applied guidelines for the selection of compounds with a greater chance of yielding successful drugs, if used in the appropriate context. However, one should be careful about being overly rigid regarding their application as this could lead to missed opportunities.

Filtering

Another simple approach to applying multiple criteria to the selection of compounds is sequential filtering. In this process the compounds are compared to series of criteria; those that fail to meet a criterion are discarded while those that meet the criterion are progressed for comparison against the next criterion in the sequence. The hope is that one or more 'ideal' compounds will emerge from the sequence of filters, having passed all of the criteria. Filtering offers the benefit that interpretation is straightforward, because if a compound fails one or more criteria this clearly indicates the focus for improving the compound.

The set of criteria against which compounds are compared can be based on any relevant properties, whether calculated or experimental. This offers the flexibility that a drug discovery project may choose criteria that are tailored to the project objectives, based on the experience of the project team or historical data for successful compounds for the intended therapeutic indication. These criteria are sometimes referred to as a target product profile (TPP) and an illustrative example of such a profile for identification of a lead compound for an orally dosed compound is shown in Table 2. Early in a project, for example when choosing a screening library for high throughput screening, it is also common to apply the criteria indicated by one of more of the rules-of-thumb discussed above as sequential filters.

Table 2. An example of a target product profile for selection of a lead compound intended for oral administration.

Property	Criterion
Pharmacology	
Potency against target (K_i)	<100 nM
Selectivity against related off-targets	>100 ×
Physicochemical	
LogP	<4
Solubility	>100 μ M
MW	<450 Da
ADME	
Caco-2* permeability (P_{app})	> 10×10^{-6} cm/s
Intrinsic Clearance in Human Liver Microsomes (Cl_{int})	<25 μ L/min/mg protein
Absence of P-glycoprotein transport (Caco2 BA:AB)	<3
Safety	
Avoid Cytochrome P450-mediated drug-drug interactions (K_i for CYP3A4, CYP2C9, CYP2D6, CYP1A2)	>1 μ M
Avoid interaction with hERG potassium ion channel (IC_{50})	>10 μ M
Cytotoxicity in HepG2 [†] cells (LD_{50})	>1 mM

*Human epithelial colorectal adenocarcinoma cell line [16]

†Hepatocellular carcinoma cell line [17]

One challenge of filtering is that it is common for no compounds to emerge from the end of the sequence; there are several possible reasons for this:

- There are often conflicts between the property criteria; improving one property often leads to an adverse change in another. In these situations, the relative importance of each criterion should be taken into account as this defines acceptable trade-offs against conflicting properties.
- Simple yes/no criteria may be too strict; For example, if a compound meets all of the criteria in the TPP, except that it has a logP of 5.1 versus a criterion of <5, does it make sense to reject it?
- There may have been a mis-measurement or mis-prediction; one or more compounds may have been incorrectly rejected due to the experimental variability or statistical error in a prediction.

The last of these is probably the biggest concern about filtering because, as we discussed above, there is significant uncertainty in almost all of the data which is available in early drug discovery. If we consider a

simple illustrative example in which we have 10 filters that are each 90% accurate in passing/failing a compound, the probability of an ideal compound emerging, even if it was present in the set being filtered, is only 35% ($p = 0.9^{10}$ assuming independence of the error in each filter). Therefore, even in this optimistic case, sequential filtering is more likely to discard an ideal compound than accept it. Furthermore, there is a significant chance of incorrectly passing a poor compound; in this example, if a compound should correctly fail only one of the criteria, the probability of it being incorrectly accepted is 4%. Given that there are typically many more poor compounds than good, this means that any ideal compound that is fortunate enough to be correctly passed by all of the filters is likely to be swamped by poor compounds incorrectly accepted.

Therefore, despite the simplicity and easy interpretation of filtering, it should be treated with caution. The process accumulates error without that being transparent, running the risk of rejecting good compounds and missing opportunities to find a high quality drug.

Calculated Metrics

Rather than defining criteria for multiple, individual properties these may be combined to calculate a single metric that can be optimized to guide selection or design. One of the earliest and most commonly applied metrics is the Ligand Efficiency (LE) proposed by Hopkins *et al.* [18], with the goal of mitigating the tendency to focus too heavily on the optimization of potency at the cost of other necessary properties. LE was derived from the observation that smaller compounds tend to have better physicochemical and ADME properties than large compounds. Therefore, given two equally potent compounds it is preferable to choose the smaller. Or, alternatively, increasing potency without significantly increasing compound size is desirable. LE is defined as:

$$LE = \frac{\Delta G}{N_H}$$

where ΔG is the free energy of binding and N_H is the number of heavy (i.e. non-Hydrogen) atoms in the compound. In more common units, this may be expressed as:

$$LE = \frac{RT \times pIC_{50}}{N_H} = \frac{1.4 \times pIC_{50}}{N_H}$$

where $pIC_{50} = -\log(IC_{50})$ and the IC_{50} is expressed in molar concentration.

The use of the LE metric is particularly popular in fragment-based drug design [19], where the starting point is typically one or more small fragments with low binding affinity and new compounds are designed by growing or linking these fragments to identify a larger compound with sufficient potency. Although the initial fragments bind only weakly, they have a high LE due to their small size and the optimization process may be guided by increasing the potency while maintaining a high LE.

The LE metric inspired other calculated optimization metrics, for example Ligand Lipophilicity Efficiency (LLE) [20], also known as Lipophilic Efficiency (LipE):

$$LLE = pIC_{50} - \log P,$$

where a calculated value of $\log P$ is often used. This was motivated by the desire to maximize potency while maintaining as low a lipophilicity as possible, due to the association between high lipophilicity and several issues including poor solubility, membrane permeation and metabolic stability, lack of selectivity and a higher risk of non-specific toxicity [21] [22].

The range of efficiency metrics has been further extended to include percent efficiency index (PEI), defined as the percent inhibition (as a fraction between 0 and 1) divided by MW in kDa; binding efficiency index (BEI), defined as pIC_{50} divided by MW in kDa; and surface efficiency index (SEI), defined as pIC_{50} divided by PSA in 100s of Å. All of these combine a measure of potency related to another property representing the 'drug-likeness' of the compound and are reviewed in detail in [23]. More complex derivatives of these efficiency indexes have also been proposed including ligand efficiency-dependent lipophilicity (LEDL), defined as $\log P$ divided by LE [24] and 'fit quality' [25].

These calculated metrics have the advantage that they are simple to apply, as only a single value must be monitored during optimization. They are also easy to interpret – Increase potency while minimizing the increase in compound size or lipophilicity – although this ease of interpretation may be sacrificed somewhat by the more complex efficiency indexes such as LEDL.

In many cases rules-of-thumb have been developed for selection of high quality compounds using these metrics. For example, it has been proposed that a LLE of 6 or higher is preferable, corresponding to a potency of better than 10 nM with a logP of 2. Again these provide useful guidelines when applied in an appropriate context, but the same caveats apply here as to the rules-of-thumb discussed above, in particular:

- Potency and logP values have significant uncertainty, particularly when predicted, yet it is rare to see the uncertainties propagated through the calculation of the efficiency metric to consider the confidence with which compounds may be chosen based on these metrics.
- As noted above, increasing compound size, MW and logP significantly increases the chance of encountering issues with poor physicochemical, ADME and toxicity. However, the correlation with these properties is not perfect, so it may be inappropriate to make selections based too strictly on these metrics, particularly when options are limited.
- These rules are not universal and are typically based on identification of orally administered drugs, so the project's therapeutic objective should be considered carefully before choosing a criterion.

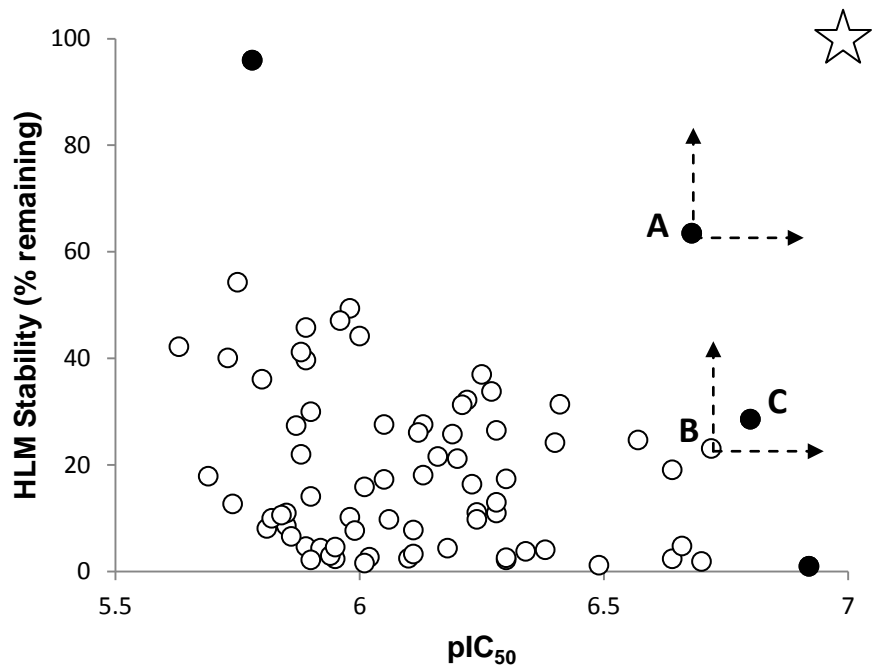
It is noteworthy that there is a close relationship between the optimization based on these metrics and a recent trend to optimize compounds based on measurements of the thermodynamic parameters of binding using biophysical measurements [26]. This strategy suggests that it is better to increase binding affinity (or equivalently decrease the free energy ΔG , as strong binding is equivalent to a reduction in free energy) by introducing an interaction dominated by decreasing the enthalpy of binding (ΔH) rather than one dominated by increasing entropy (ΔS) – note $\Delta G = \Delta H - T\Delta S$, where T is the temperature. Decreasing the binding free energy by reducing the enthalpy is achieved by forming a specific interaction with the target, for example a hydrogen bond with a residue in the binding pocket, which will typically improve the LE or LLE. The free energy can also be reduced by increasing the entropy and this can be achieved by displacing coordinated water molecules from the binding pocket into bulk solvent, for example by adding a bulky lipophilic group to occupy the binding pocket. However, this is often detrimental in the long-run, as such a non-specific interaction will increase the chance of off-target binding or non-specific toxicity and this is reflected by a decrease in the LE or LLE [27].

Pareto Optimization

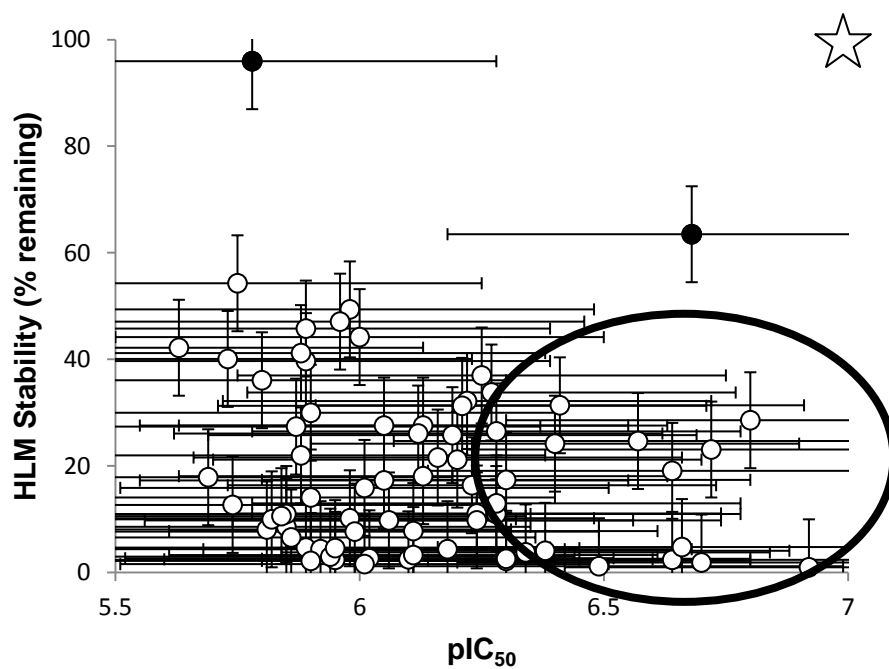
The concept known as Pareto optimality was proposed by an Italian economist Vilfredo Pareto in the early 20th Century [28]. He suggested that, when considering multiple parameters, there may not be a single best combination of parameters, but rather a family of solutions that each represents a different, optimal combination. More specifically, a solution to a multi-parameter optimization problem is considered to be a Pareto optimum if there is no other solution that is better in all of the parameters.

To illustrate this, consider the two-parameter example, illustrated in Figure 2(a), of a hypothetical drug discovery project which wishes to achieve an optimal balance of potency and metabolic stability to achieve good *in vivo* exposure and hence efficacy. Ideally, the project would like to identify a compound with high potency (pIC_{50}) and good stability in human liver microsomes (% remaining after incubation for 40 minutes). This ideal goal is represented by the top right corner of the plots in Figure 2. However, as this ideal may be difficult or impossible to achieve, the project would like to find a good balance between potency and metabolic stability. The points shown as solid points in Figure 2 represent compounds with different, Pareto optimal combinations of these two parameters. For example, the point labeled A has no points that have both better potency and better metabolic stability, i.e. there are no points to the right and above; such a point is described as 'non-dominated'. Contrast this with the point labeled B, which has a point, C, to the right and above representing a compound that is better in both parameters; B is 'dominated' by C. Note that the non-dominated points define a boundary, known as the 'Pareto front' and each represents a candidate for further investigation to identify the best balance of potency and metabolic stability to achieve *in vivo* efficacy.

The concept of Pareto optimality may be generalized to Pareto rank, whereby a point is ranked according to the number of points by which it is dominated, a rank-0 point is non-dominated, rank-1 is dominated by only a single point, etc. This allows compounds to be ranked according to how close they are to the optimum front.



(a)



(b)

Figure 2. Plots illustrating the concept of Pareto optimality in which the ideal outcome corresponds to the top right corner of the plot, as indicated by the star: (a) shows a scatter plot of data for activity against the therapeutic target and stability in human liver microsomes for a set of 75 compounds. The solid points are Pareto optimal or 'non-dominated' points; for example, in the case of point A, there are no points with a higher value for both parameters. However, open circles are not Pareto optimal; for example point B is 'dominated' by point C. (b) shows the same data for which the uncertainty (1 standard deviation) has been shown as error bars on each point. From this it is clear that while two points may be confidently identified as Pareto optimal (solid circles) there are many points in the region indicated by the ellipse, which may not be confidently identified as Pareto optimal or otherwise.

Pareto optimization has a strong benefit of offering flexibility if the appropriate weighting for each property is not known *a priori* or when it is useful to explore a range of different potential solutions. It is also possible to apply a weighting toward one or more properties by selecting compounds from only a portion of the Pareto front. Furthermore, interpretation is quite straightforward, at least for small numbers of parameters, as these solutions can be easily visualized.

However, when dealing with a large number of parameters (e.g. 5 or more) to be optimized, Pareto optimization becomes less useful because the number of solutions on the Pareto front grows exponentially with the number of parameters. This leads to an overwhelming number of 'optimal' solutions making it impossible to evaluate them all and choose between them. Unfortunately, in the drug discovery process, it is common to deal with many more than 5 parameters. One approach to addressing this is to combine multiple individual parameters into a small number of 'scores' representing different factors that are then subject to Pareto optimization. For example, all of the parameters relating to ADME properties may be combined into a single ADME score and the trade-off explored with potency using Pareto optimization [29]. The integration of individual properties into a single score may be achieved using a method such as desirability functions or probabilistic scoring, as described below.

A further limitation of the Pareto optimization approach is that it does not explicitly take the uncertainty of the underlying data into consideration. For example, Figure 2(b) shows the same data for the compounds in Figure 2(a) with the uncertainty shown by error bars for each point. From this it can be seen that some compounds can be confidently identified as Pareto optimal; however in many cases it cannot be confidently determined which compounds are Pareto optimal, suggesting a number of compounds that are not on the explicit Pareto front are worthy of consideration.

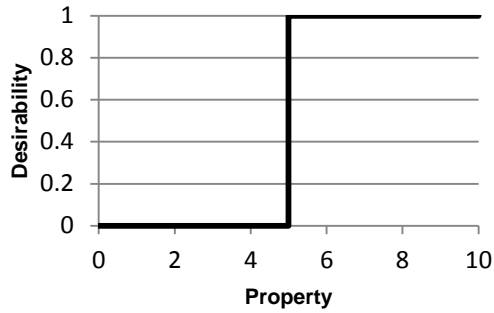
Therefore, for a small number of parameters to be optimized, Pareto optimization provides an excellent approach to investigate the best trade-off between the competing factors. However, for data with high dimensionality or uncertainty, the number of potential solutions becomes too large for easy consideration and the Pareto approach must be combined with another method to reduce the complexity of the data.

Desirability Functions

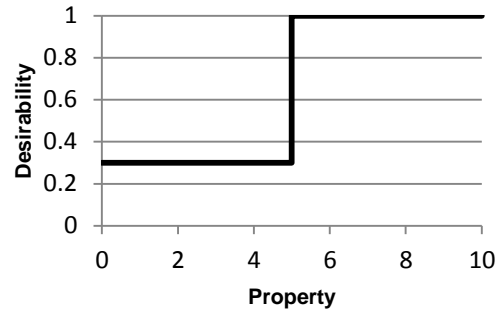
The desirability function was first proposed by Harrington [30] in 1965 as an approach for combining multiple responses in a single optimization equation. A desirability function maps the value of a property onto a score in the range from zero to one that represents how desirable a compound with this property value would be. If the property of a compound lies in the ideal range, it will be given a score of 1.0 and if the property is such that the compound would be absolutely rejected it would be given a score of 0.0. Scores between zero and one represent increasing levels of desirability.

A simple pass/fail filter may be defined easily using a desirability function that has a value of 1.0 on the desirable side of the criterion value and a value of 0.0 on the undesirable side, as illustrated in Figure 3(a). However, desirability functions provide great flexibility in defining the criteria and importance of a property for the identification of high quality compounds. A less important property would receive a score greater than zero on the undesirable side, indicating that this 'failure' would not be cause for rejecting the compound absolutely, as illustrated in Figure 3(b). The importance of the property to determining the quality of a compound is reflected by the minimum desirability score it can achieve; a critical property is one for which an unacceptable value would lead to the outright rejection of a compound.

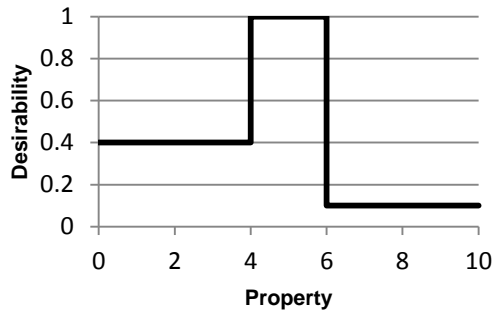
Other shapes of function can define an optimal range (Figure 3(c)), a single optimal value (Figure 3(d)) or a trend across a range of property values from absolute rejection to ideal (Figure 3(e)). The functions in Figure 3(a-e) are all examples of linear forms of a class of functions called Derringer desirability functions [31] which also include non-linear variants that can reflect the speed at which the desirability decreases as the property moves away from the ideal value (an example is shown in Figure 3(f)).



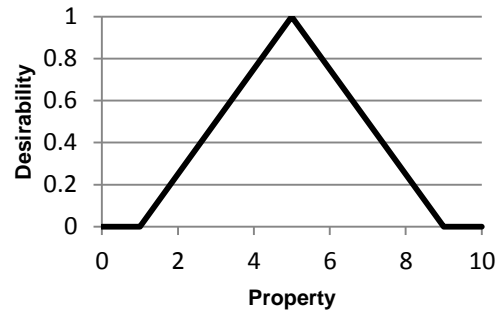
(a)



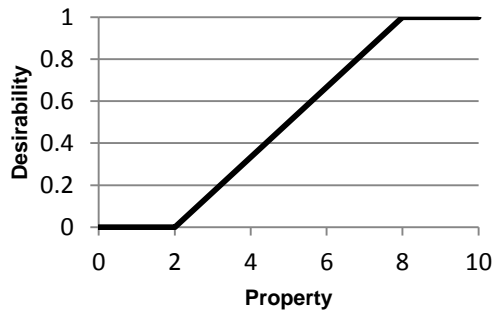
(b)



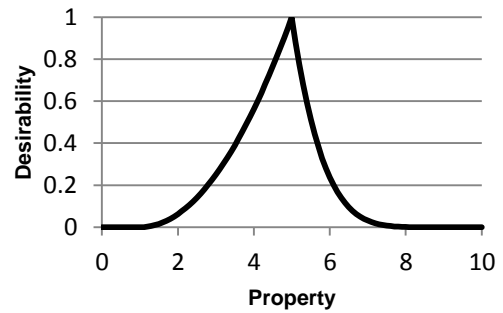
(c)



(d)



(e)



(f)

Figure 3. Example desirability functions. The functions shown by the bold lines in (a) through (e) are examples of linear desirability functions: (a) is a threshold function representing a simple filter, as property values that do not achieve the criterion of >5 have a desirability score of zero; (b) is a threshold function with less importance, as property values that do not achieve the criterion of >5 are less desirable, but will not be rejected outright; (c) defines an ideal property range of 4-6 with values exceeding the upper limit being less desirable than those less than the lower limit; (d) defines an ideal value of 5 with linearly decreasing desirability above and below this and (e) defines an ideal property criterion of >8 with linearly increasing desirability above a value of 2. The desirability function in (f) is an example of a nonlinear Derringer function with desirability d for property value Y defined by:

$$\begin{aligned}
 Y < 1: & \quad d = 0 \\
 1 < Y < 5: & \quad d = \left(\frac{Y-1}{4}\right)^2 \\
 5 < Y < 8: & \quad d = \left(\frac{9-Y}{4}\right)^5 \\
 8 < Y: & \quad d = 0
 \end{aligned}$$

A single desirability function only provides a way to assess the quality of a compound according to one property. However, by mapping all properties onto a desirability scale between zero and one, the individual desirability scores due to multiple properties may be easily combined, even if the properties have different scales or units of measurement. Multiple desirability scores corresponding to different properties can be combined into a 'desirability index', which is a measure of the overall quality of a compound.

There are two common approaches to defining a desirability index: an additive approach combines the individual desirability scores by taking the average,

$$D = \frac{d_1(Y_1) + d_2(Y_2) + \dots + d_n(Y_n)}{n},$$

where D is the overall desirability index, $d_i(Y_i)$ is the desirability score for property Y_i and n is the number of properties. Alternatively, a multiplicative approach takes the geometric mean of the individual desirability scores;

$$D = (d_1(Y_1) \times d_2(Y_2) \times \dots \times d_n(Y_n))^{1/n}.$$

A disadvantage of an additive approach is that if a large number of properties are being combined in an assessment of the overall desirability a very low desirability for a single property will only have a small impact on the desirability. However, if a compound has an unacceptable value of a critical property the compound should be rejected, e.g. a compound with 100 μM IC_{50} is not of interest even if it has ideal ADME properties. This behavior is captured by a multiplicative definition of the desirability index.

As the desirability index is calculated from the individual desirability scores corresponding to each property, the impact of each property on the overall quality of a compound can be easily identified. This provides for easy interpretation of the overall result and quickly identifies the most important issues that should be addressed in order to optimize a compound to improve its overall quality.

A limitation of the desirability function approach is that it assumes an *a priori* knowledge of the trade-offs between the different factors contributing to the success of a compound against the therapeutic goal of a project. This relies on the expert domain knowledge of the project team, which introduces a degree of subjectivity into the prioritization of criteria and hence compounds. However, this approach also provides a straightforward way to test the effect of this subjectivity, by exploring the impact of changes in the desirability functions on the selection of compounds. Thus, it can help to challenge the decision-making process and focus attention on critical experiments, e.g. *in vivo* studies, which will help to identify the most appropriate profile to select high quality compounds with greater accuracy [32].

Therefore, desirability functions provide a flexible method to define an ideal property profile and assign a weight for each property criterion to prioritize compounds in an easily interpretable manner. Defining 'soft' boundaries to a desirability function, rather than a hard cut-off helps to avoid rejection of compounds based on an uncertain property value close to a criterion boundary. However, the standard desirability function approach does not explicitly consider the uncertainty in the underlying data and in the next section we will describe an extension of this method that adds the ability to assess the confidence with which compounds may be distinguished.

Probabilistic Scoring

The probabilistic scoring approach [33] [34] builds on desirability functions by explicitly incorporating the uncertainty of the underlying data to provide confidence and objectivity in decisions regarding compound selection.

The Importance of Uncertainty

In order to illustrate the importance of uncertainty in selection of compounds, consider the examples shown in Figure 4. Figure 4(a) shows a desirability function corresponding to a simple filter, which will accept compounds with values of property value less than 5 and reject those with property value greater than this. Compounds A, B, and C have property values 4, 6 and 8 respectively, therefore B and C would receive a desirability score of 0.0 and be rejected, while compound A would receive a desirability score of 1.0 and be accepted. However, if the property values are uncertain, as illustrated by Gaussian probability distributions,

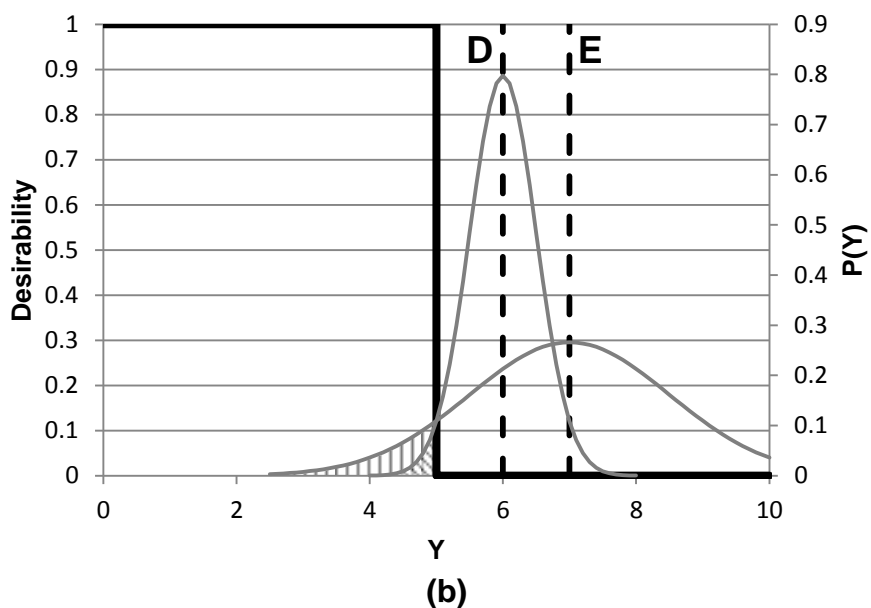
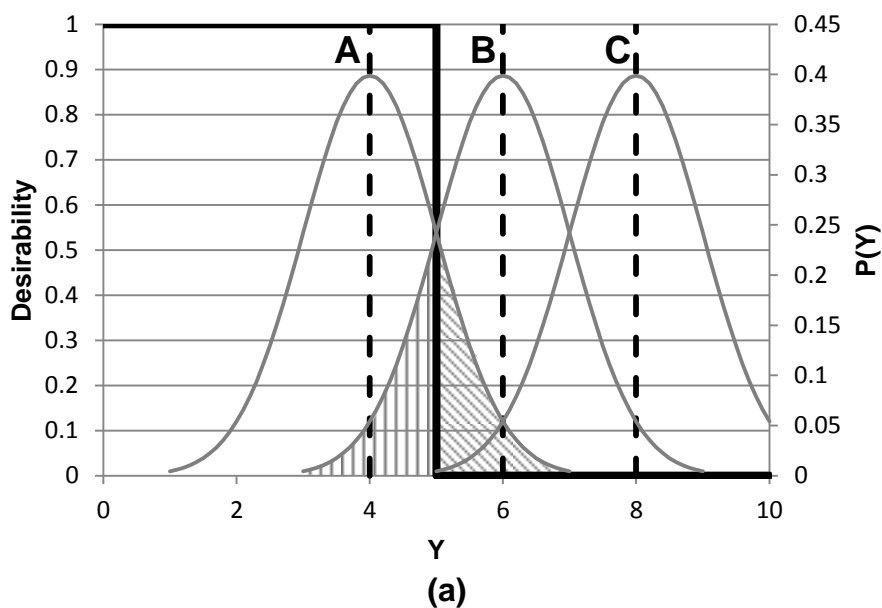


Figure 4. Illustrations of the importance of uncertainty when selecting compounds. Both figures show a desirability function (bold line) corresponding to a simple filter with a criterion of <4 . The dashed vertical lines indicate values of property Y for compounds labeled A through E. The uncertainties in these property values are illustrated by the grey bell curves (Gaussian distributions) centered on each compound's property value. In (a) if we were to ignore the uncertainties in the property values, compound A would be accepted and B and C would be rejected. However, considering the uncertainties, we can see that, while the probability of compound C achieving the criterion is negligible, there is a significant probability, shown by the vertically hatched area, that compound B will meet the criterion and there is an equal probability, shown by the diagonally hatched area, that compound A will not meet the criterion. In (b) we can see that the values for both compounds D and E fail to meet the criterion. However, taking the uncertainties into account we can see that, even though the value for D is closer to the criterion than E, the probability of compound E meeting the criterion (the vertically hatched area) is actually greater than that for compound D (the diagonally hatched area).

we can see that this decision should not be so clear-cut. While we can see that compound C is very unlikely to meet the criterion, there is a significant probability that compound B will achieve the required criterion, while there also a significant probability that compound A will not achieve the criterion. Therefore, a more appropriate interpretation of this data is that C can be rejected with confidence, but compounds A and B cannot be confidently distinguished. Better data or another criterion would be needed to select between A and B.

The example shown in Figure 4(b) illustrates another subtle effect of uncertainty. The compounds D and E both have property values on the undesirable side of the criterion and would receive a desirability score of 0.0. However, if these were the only two options and one must be chosen for further testing, it may be tempting to select compound D, as it is closer to the criterion value. However, it should be remembered that the same properties can be measured to different levels of accuracy and so the uncertainties in the data, even for a single property, will not necessarily be the same. As such, if we consider the uncertainties in these values, again illustrated as Gaussian distributions, we can see that the value of compound E is less certain and hence the probability of compound E achieving an acceptable value is higher. Therefore, it would be more appropriate to select compound E; it is better to select a compound we are uncertain about, instead of a compound we are confident will fail. Of course, after selecting compound E a more accurate value for the property should be determined as a priority.

Combining Desirability with Uncertainty

Probabilistic scoring [33] [34] allows the project to define the property profile required for a successful compound (a 'scoring profile') using desirability functions to provide flexibility in defining the individual property criteria and their importance to the overall objective of the project. The question of desirability is separated from the confidence in the property values produced by an experiment or prediction; the desirability can be defined under the assumption that the data on a compound's properties could be determined perfectly. An example of such a scoring profile is shown in Figure 5, which combines experimental potency with predicted ADMET properties.

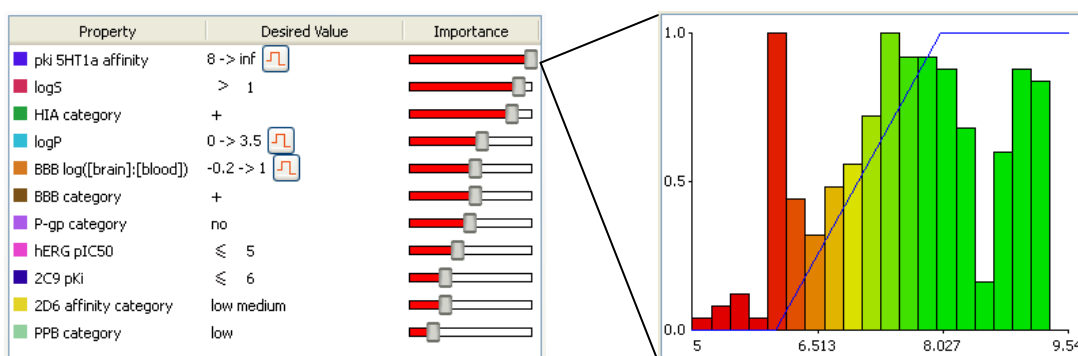


Figure 5. An example of a scoring profile. For each property the ideal outcome (desired value) and importance is indicated. Underlying each property criterion is a desirability function as illustrated in this case for the pK_i against the 5HT_{1a} target. The desirability function (bold line) is superimposed on a histogram showing the distribution of the underlying data.

The data for each compound is then assessed against the scoring profile, to calculate a probabilistic score for each compound, which represents the most likely desirability score for the compound, taking into account the uncertainty in each data point. The uncertainty in the overall score provides valuable additional information that helps to make decisions with confidence; it indicates clearly when compounds may be confidently distinguished given the uncertainty in the underlying data.

The uncertainty in the data may derive from variability in an experimental assay or the statistical error in a property prediction and may vary from compound to compound. For example, when multiple experimental replicates have been performed, the best estimate of the property value and uncertainty are the mean and standard error in the mean of the individual measurements, which will be different for each compound. In cases where data is missing for a compound, e.g. when an assay has not yet been performed, this can also be treated rigorously, as a missing piece of data is simply a data point with a very high uncertainty. Therefore, if a compound meets all criteria with high confidence, except for an important property for which a data point is missing, the result would be a moderate score, with very high uncertainty, i.e. if the missing data point were measured and the result was good, this would be an ideal compound, but if the measured value were poor,

the compound would not be acceptable. Whereas, if a compound fails multiple criteria with high confidence but has a missing data point, the score will be low with low uncertainty, as the missing data point is unimportant; a good result for one property would not 'rescue' a poor compound.

One illustration of the output is shown in Figure 6, which shows the calculated probabilistic scores for a set of compounds against the profile shown in Figure 5. From this it is clear that, although the first compound achieves the highest score, the uncertainty in the score for this compound means that it cannot be confidently selected over several of the top-scoring compounds; the error bars on the scores overlap significantly, suggesting that more precise data or another criterion will be necessary to choose between these compounds. Conversely, we can see that approximately half of the compounds may be rejected with confidence, as the error bars for these compounds do not overlap with those of the top-scoring compounds, indicating that the chance of them being of similar quality to the best compounds in the set is negligible.

Interpretation of probabilistic scores is also straightforward, as the contribution of each property to the overall score may be identified, giving clear guidance on the focus of further optimization efforts. One approach to visualizing this is a histogram showing the impact of each property on a compound's score, as illustrated inset in Figure 6. The height of each bar corresponds to the contribution of the property to the score; a very low bar indicates a property that fails to meet the criterion for an important property with high confidence, while a high bar indicates a good result for a property with high confidence. Other ways this information could be displayed include a heat map or radar plot, as illustrated in Figure 7.

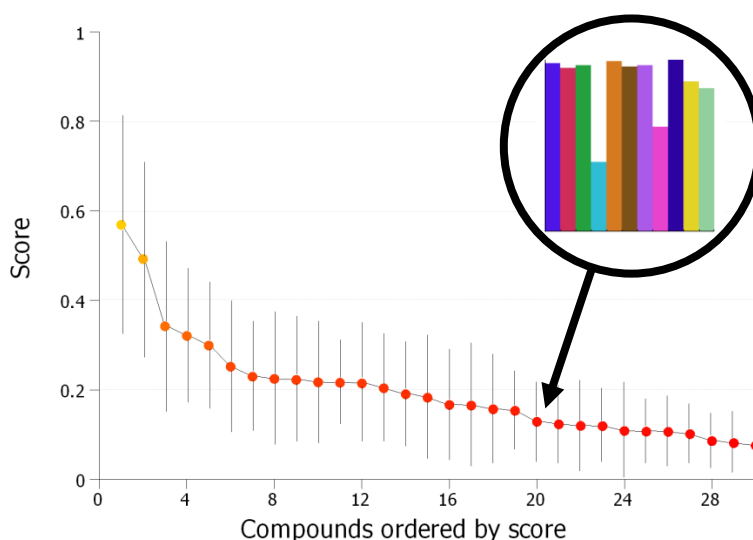


Figure 6. An example output from probabilistic scoring for 30 compounds. The compounds are ordered from left to right along the x-axis in order of their score and overall score for each compound is plotted on the y-axis. The overall uncertainty in each score (1 standard deviation), due to the uncertainty in the underlying data, is shown by error bars around the corresponding point. The impact of each individual property on the overall score can be interpreted and an example histogram visualizing this for a single compound is shown inset. High bars indicate properties that achieve the ideal outcome with high probability while low bars indicate properties that have a significant negative impact on the score. The colors of the bars correspond to the key in the profile shown in Figure 5.

As probabilistic scoring is based on the foundation of desirability functions, the same limitation applies, namely that an *a priori* knowledge of an appropriate property profile is assumed. Furthermore, an understanding of the uncertainty in the underlying property data, whether due to statistical error in predictions or experimental variability, is required. Ideally, this should be determined as part of the validation of a predictive model or assay. Where multiple experimental replicates are available, the uncertainty can be estimated for each compound as the standard error in the mean of the individual measurements. However, even when a quantitative measure of the uncertainty is not available, it may still be valuable to include an estimate of the confidence based on the experience of the experimentalist, who will often have a feel for how much they 'trust' the data. In this way, a scientist without this domain knowledge can take into account the expert's interpretation of the data and avoid giving too much weight to differences in property values that are not significant.

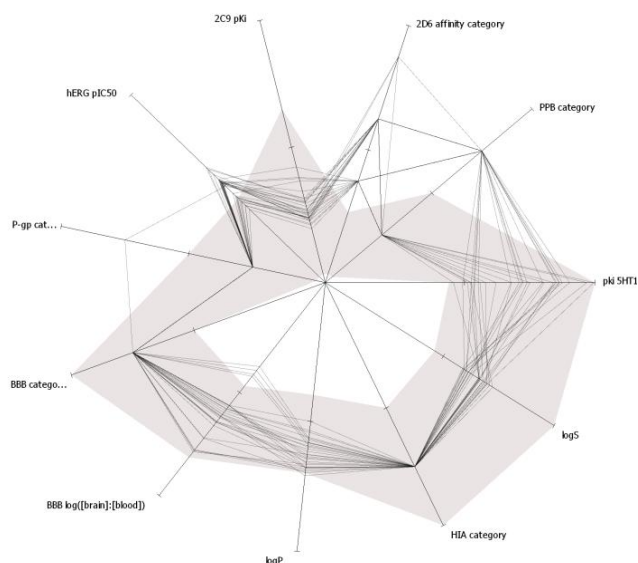


Figure 7. An example radar plot visualization for probabilistic scores for a set of compounds. Each radial axis represents a single property in the scoring profile shown in Figure 5. The grey region shows the ideal range for each property and each compound is illustrated by a line joining the axes, such that the intercept indicates the value of the corresponding property. This can help to visualize the performance of the overall set relative to the ideal property profile and each individual compound relative to both the ideal profile and overall set.

Therefore, this approach brings together the requirements outlined in the Introduction to allow flexibility in setting property criteria for each project and the weight given to each individual property criterion, while taking the uncertainty in the available data into account and maintaining easy interpretation of the results.

Multi-Objective Evolutionary Algorithms

In some cases, the number of possibilities to prioritize is too high to simply enumerate and rank them all using one of the methods described above. Searching ‘chemical space’ is particularly challenging, as it is not a continuous space, i.e. there are only certain discrete changes that can be made to a compound. Nor are the responses of biological properties to chemical structure smooth; there are numerous anecdotes of the ‘magic methyl’ whereby a small change in the right place on a compound dramatically changes its properties [35]. These factors, combined with the enormous size of the potential search space – the number of possible ‘drug-like’ compounds has been estimated to be $\sim 10^{60}$ [36] – means that a computational algorithm must be used to explore the ‘space’ of possibilities in search of high quality solutions.

A popular class of algorithms for searching large, complex spaces are ‘evolutionary algorithms’ (EAs) [37], which are motivated by the theory of evolution. These algorithms ‘evolve’ a population of potential solutions, ‘combining’ and randomly ‘mutating’ their features before selecting the ‘fittest’ as the basis for the subsequent ‘generation’. These stochastic methods cannot be guaranteed to find the optimal solution, but used appropriately, can find good solutions with a high probability.

De novo design algorithms generate new compound structures, often using an EA approach. The output of such an algorithm can then be prioritized using an MPO approach to find the ‘fittest’, i.e. the highest scoring, as the basis for the next generation. In an evolutionary approach, new structures may be generated by ‘mutating’ features of a compound (e.g. changing bond orders, breaking or forming rings, introducing heteroatoms or substituting new atoms) or combining fragments from multiple compounds to create a new, hybrid compound [38]. An alternative method is to use ‘medicinal chemistry rules’ for modifying compounds, taken from the experience of medicinal chemists, to define compound transformations that may be applied iteratively to create new ‘generations’ of compounds [39]. In these ways, *de novo* design, coupled with MPO, can be used as a directed search for high quality compounds that are likely to meet the criteria for a project [40] [41] and suggest new structures that may be interesting for further investigation. One of the limitations of *de novo* design is the potential to propose unstable, unfeasible or synthetically intractable chemical structures. One of the advantages of the transformation rule based methods is that they tend to produce more acceptable structures, due to the fact that the transformations are based on historical precedents. However,

computational methods to estimate the synthetic tractability of compounds can help to discard irrelevant compounds generated by inappropriate mutations or combinations of structures [42].

Another class of problems which often yields an unfeasibly large number of possibilities to sample exhaustively is the selection of a subset of compounds from a large (virtual) library. For example, even when only selecting 10 compounds from a possible 100 there are 1.7×10^{13} different combinations. Of course, if there is a free choice of compounds to select, the obvious solution is simply to pick the highest scoring compounds. However, in some cases it may be necessary to restrict the selection of compounds to meet constraints for synthesis, e.g. a combinatorial selection to allow parallel synthesis or limit the number of reagents required. Even when a free choice is available, simply selecting the highest scoring compounds may not be the best approach, as the top scoring compounds may be very similar in terms of their chemical structure. In this case, it may be appropriate to temper a focus on score with an element of structural diversity to explore a range of different potential chemistries. Structural diversity is a property of a collection of compounds and individual compounds cannot be assigned a measure of their diversity. Therefore, a large number of potential selections must be explored in order to search for an appropriate degree of diversity. Fortunately, this library design problem is well suited for an EA approach as discussed in more detail in the following section.

A survey of many algorithms for molecular optimization using computational MPO algorithms is available in reference [43]

Library Design: Balancing Quality and Diversity

As discussed above, when selecting sets of compounds from a larger library, it is often valuable to not focus exclusively on the 'best' compounds. As we have discussed, the available data is usually uncertain and closely related compounds are more likely to share a common cause of unpredictable failure at a later stage (e.g. a common mechanism of toxicity). Therefore, it is often useful to balance quality with an exploration of structural diversity to validate predicted hypotheses, mitigating risk by exploring potential backup series and gathering information on structure-activity relationships.

There are many definitions of structural diversity that may be used in this context. The difference between pairs of compounds may be considered in terms of the two-dimensional structure (the graph of atoms and bonds), for example using fingerprints based on sub-structural keys or atom paths and a Tanimoto similarity index [44]. Alternatively, methods that take three-dimensional conformational or shape-based information may be used. There are also many ways to assess the diversity of a set of compounds, including the minimum or average difference between compounds in the set using one of the pair-wise difference measures, clustering of compounds into structurally similar groups or dividing a descriptor space into 'cells' and sampling compounds evenly from each cell. An excellent overview of methods for assessing the diversity of compounds and libraries is provided in [45] and the references therein.

The problem of library design and selection maps neatly onto a class of EAs called Genetic Algorithms (GAs) [46]. In a GA, the characteristics of a member of the population of potential solutions is represented using a 'genetic code' which can then be 'mutated' or 'crossed' with the genetic code of another member to create 'offspring', mimicking the transfer of genetic material during reproduction. For selection of compounds, the genetic code can be a binary string in which the presence or absence of a compound or, in the case of a combinatorial library a reagent group, is encoded as a 1 or 0 respectively. Alternatively the genetic code can be represented as a list of integers representing the selected compounds or reagent groups from the available pool. A mutation corresponds to changing a 0 to a 1 or vice versa or changing one integer value to another. Crossing two members of the population involves splitting the codes, swapping a section of each and recombining the sections. In both cases, care must be taken to ensure that the appropriate number of compounds or reagents is selected in total.

One example of such an approach is the MoSELECT algorithm published by Gillet *et al.* [47] which combines a genetic algorithm with Pareto optimization to evolve combinatorial library design strategies using the Multi-objective Genetic Algorithm (MoGA) [48]. This approach explores different strategies for reagent selection and, in each generation, chooses the library designs as the basis for the next generation based on their Pareto rank. The Pareto rank is calculated for the properties of the library to be optimized, including diversity, cost of reagents, and measures of 'drug-likeness' such as MW, HBD and HBA. The algorithm evolves increasingly good approximations to the library designs that lie on the Pareto front and the result is a family of possible library

designs that range from highly diverse to those focused on one of the property criteria and cover combinations between these extremes. As previously discussed, this has the advantage of providing a number of possible solutions for the project to explore, which is particularly useful if the best balance of diversity versus other quality criteria is not known *a priori*. However, the choice can become overwhelming, particularly if there are no simple *post-hoc* criteria with which to distinguish the different library designs.

Another approach to balancing quality and diversity is to combine a measure of library diversity with a measure of quality into a single optimization metric. Most commonly, this is achieved by defining a fitness function as the weighted sum of diversity and functions of other properties of the selected set, allowing the user to choose an appropriate balance between quality and diversity, i.e.

$$F = w_d D + w_1 f_1(p_1) + w_2 f_2(p_2) + \dots,$$

where w_d is the weight given to diversity, D is the diversity of the selected set, w_i is the weight given to property i and $f_i(p_i)$ is a function of the property value p_i , for example the average of the values of a desirability function for the selected compounds. Typically, the weights are normalized such that they sum to one. This may be simplified further into a weighted sum of diversity and a score, for example a desirability index or probabilistic score, which combines the non-diversity quality metrics, i.e.

$$F = w_d D + w_s S,$$

where w_s is the weight assigned to the score and S is the score for the selected set, e.g. the average score for the selected compounds. Maximization of such a two-component fitness function is equivalent to searching for the Pareto front along a single direction in the (D,S) space, as illustrated in Figure 8. Setting $w_d = 1$ and $w_s = 0$ will search for the most diverse possible selection and the opposite values will search for the highest scoring selection. Examples of the results of this approach for free selection of compounds from a screening library are shown in Figure 9. In this example, a probabilistic score was used to assess the quality of the compounds against the project's objectives and different weights were considered for score and diversity. From this, one can see that, given the uncertainties in the overall scores, it is possible to explore significant additional diversity without a statistically significant sacrifice in compound quality. A further example of this approach for combinatorial library design may be found in reference [33].

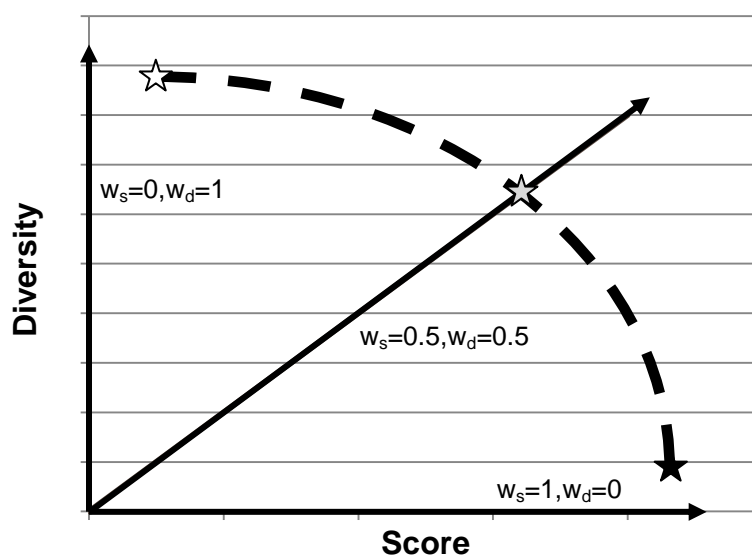
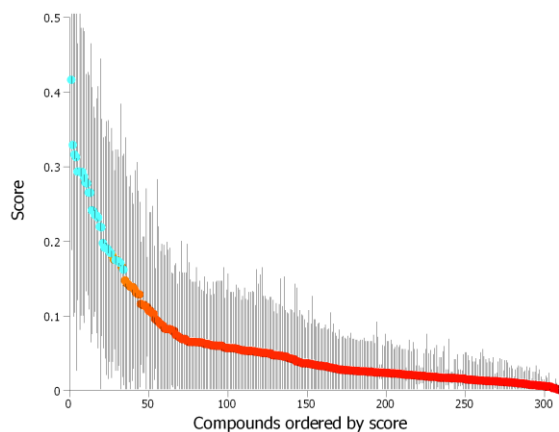
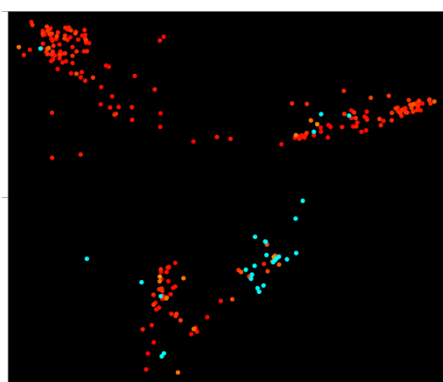
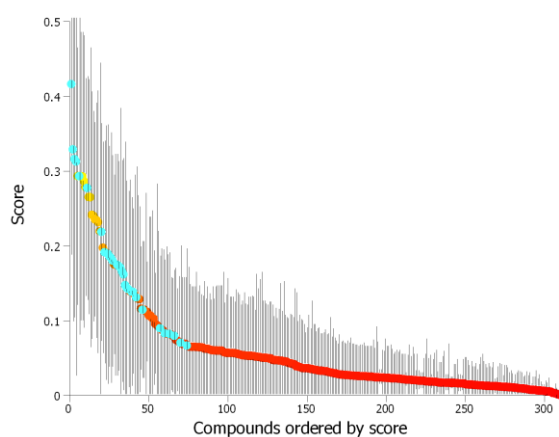
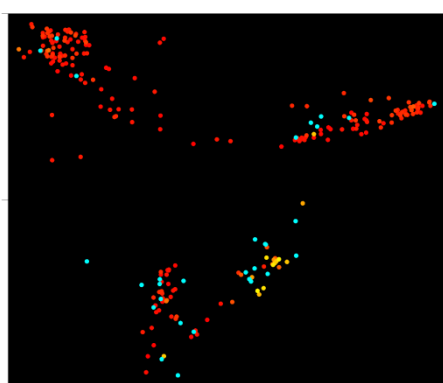


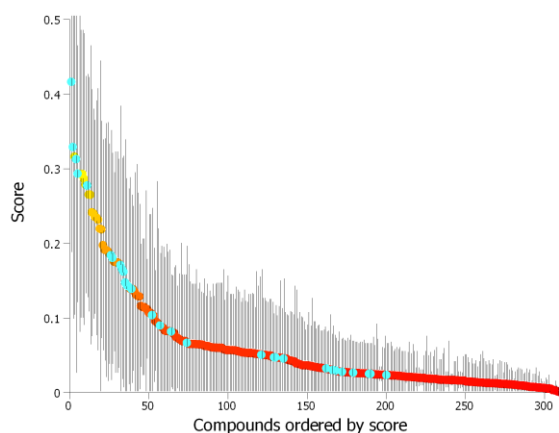
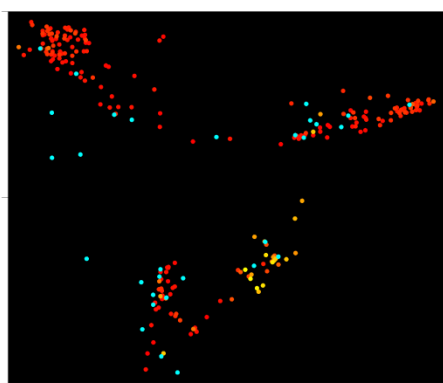
Figure 8. Diagram illustrating the relationship between Pareto optimization of score versus diversity and optimization along a single direction defined by a fitness function corresponding to a weighted sum of score and diversity with weights w_s and w_d respectively. The Pareto front of optimal selections with different balances of score and diversity is indicated by the dashed line. The search directions corresponding to three different fitness functions are indicated by arrows annotated with the corresponding values of w_s and w_d . Searching along the direction $w_s=1, w_d=0$ will identify the highest scoring selection indicated by the black star; searching along the $w_s=0.5, w_d=0.5$ direction will yield a selection with an equal balance of score and diversity, as indicated by the grey star and searching along the $w_s=0, w_d=1$ direction will identify the most diverse selection indicated by the white star.



$w_s=0.8, w_d=0.2$



$w_s=0.5, w_d=0.5$



$w_s=0.2, w_d=0.8$

Figure 9. Three example selections corresponding to different balances of score and diversity from a library of compounds scored using probabilistic scoring. The plot on the left for each selection is a 'chemical space' plot illustrating the structural diversity of the library. These are plots of the first two principal components of the similarity space calculated using 2-dimensional, path-based fingerprints and a Tanimoto similarity index. On the right is a scoring plot for the compounds in the library, similar to that in Figure 6. The points are colored from high score (yellow) to low score (red) and the selected compounds are colored in light blue. From this we can see that the selection corresponding to $w_s=0.8, w_d=0.2$ focuses on the highest scoring compounds, but samples a relatively limited region of chemical space. For $w_s=0.5, w_d=0.5$ compounds with a broader range of scores (but still within the highest ~25% of scores) and a wider diversity are selected. Finally, for $w_s=0.2, w_d=0.8$ the selection results in sampling compounds across greater diversity and all but the lowest scores.

The advantage of this approach is that, if a small number of possible trade-offs for quality versus diversity are considered interesting *a priori*, these may be efficiently explored under complete control by the user. However, if the most relevant trade-offs are not known, the Pareto optimization approach allows a more comprehensive exploration of the possibilities.

Example Applications

We have reviewed several methods for MPO in the context of drug discovery. In this section we will present illustrative examples of the application of desirability functions, Pareto optimization and probabilistic scoring to practical drug discovery challenges.

Desirability Functions for MPO of Central Nervous System Drugs

Wager *et al.* [49] presented the derivation of an MPO system, based on desirability functions, for the identification of compounds with a higher probability of success against a central nervous system (CNS) target. The authors focused on six physicochemical parameters: clogP, clogD, MW, TPSA, HBD and the pKa of the most basic center. Desirability functions were derived for these properties as shown in Figure 10.

In this implementation, the desirability score for each property is summed to give a desirability index between 0 and 6. The authors compared the desirability indexes for a set of 119 marketed drugs for CNS targets with 108 Pfizer CNS candidates and found that 74% of the marketed set achieved a desirability score of ≥ 4 compared with only 60% of the Pfizer candidates, a statistically significant difference.

The authors also explored the relationship of the desirability index with key *in vitro* ADME and safety endpoints, specifically: membrane permeability (apparent permeability (P_{app}) measured in the Madin-Darby canine kidney (MDCK) cell line), P-glycoprotein (P-gp) efflux liability (measured in MDCK cells transfected with the MDR1 gene), metabolic stability in human liver microsomes (HLM) (unbound intrinsic clearance), and general cellular toxicity (measured in a THLE Cv assay). For each of these endpoints, compounds with a desirability index of >5 had significantly higher odds of achieving a favorable outcome in the assay than compounds with a low desirability index. This was found both for the marketed drug and candidate sets.

Finally, a larger set of 11,303 Pfizer compounds were studied and the desirability indexes compared with results of the *in vitro* MDCK, P-gp and HLM assays described above, plus a dofetilide binding assay which is an indicator of risk of interaction with the potassium channel encoded by the human ether-a-go-go related gene (hERG). Again, a high desirability index was found to correspond to a significantly higher probability of achieving a favorable outcome for each of these assays. Furthermore, compounds with a high desirability index were found to have a greater chance of achieving favorable results in all four of the *in vitro* endpoints simultaneously, the ideal outcome.

Therefore, the authors concluded that the CNS MPO scheme would be useful for evaluating design ideas, triaging high-throughput screening hits and prioritizing compounds with a higher probability of successfully testing hypotheses in the clinic. They also emphasized the value of the flexibility in design provided by the use of desirability functions over 'hard' cut-offs (i.e. filters).

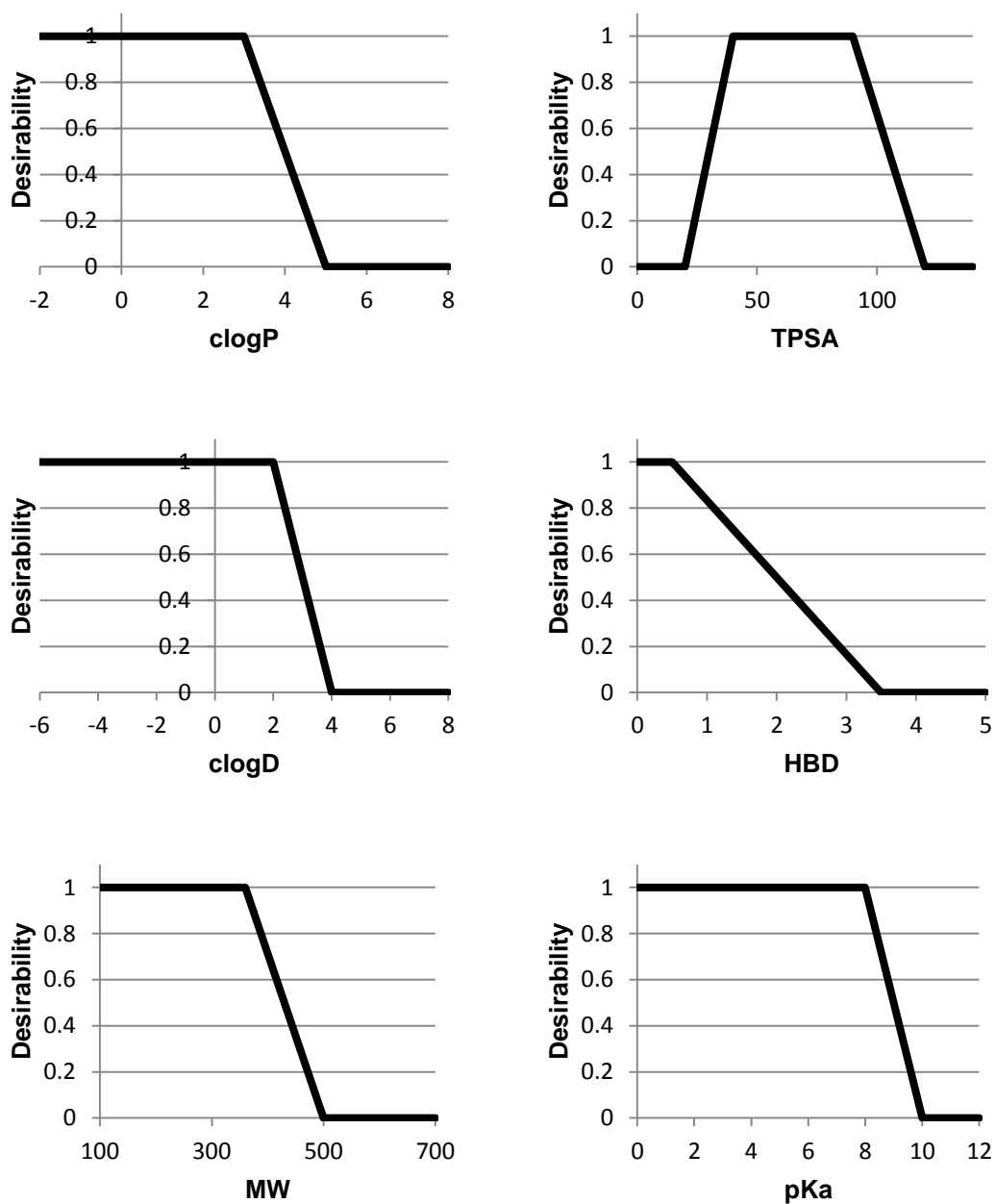


Figure 10. The desirability functions defined in Wager *et al.* [49] for selection of compounds intended for a CNS target. Desirability functions are defined for calculated logP (clogP), calculated logD at pH=7.4 (clogD), molecular weight (MW), topological polar surface area (TPSA), the number of hydrogen bond donors (HBD) and the pKa of the most basic site.

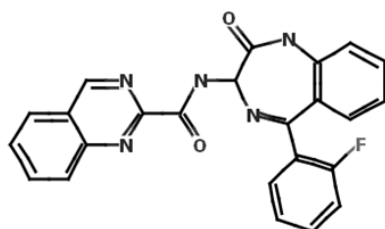
Evolving Molecules using Pareto Optimization

Ekins *et al.* described a software tool, "Pareto Ligand Designer," that combines molecular transformation with Pareto optimization to evolve initial compounds with sub-optimal properties toward structures that are more likely to achieve a required property profile [50]. The algorithm employs an engine that transforms compounds using a set of rules similar to those implemented in the Drug Guru package [39]. The compounds generated are subjected to series of filters corresponding to desired values of predicted properties and an absence of structure alerts and then the Pareto optimal compounds are identified and saved before the remaining compounds are used as the input for the next iteration.

In one example application, Pareto Ligand Designer was applied to a known CCK antagonist reported by Evans *et al.* [51] and shown in Figure 11(a). This compound has good biological activity, but is predicted to have poor blood-brain barrier (BBB) penetration and poor aqueous solubility. The objective in this case was to identify active compounds with improved predicted BBB and solubility, which were the parameters chosen for Pareto optimization. In addition, to ensure that the potential for activity against the target and other 'drug like' properties were retained in the compounds generated, a series of filters were applied at each iteration: a minimum similarity to the initial compound of 0.35, using ECFP_6 fingerprints [52] and Tanimoto similarity index [44], which corresponded to an activity belief of 16.6% using Belief Theory [53]; MW < 500 Da, 0.0 < clogP < 5.0; and an absence of alerts for undesirable substructures [54].

The authors observed that the mean objective function for the compounds selected at each iteration decreased dramatically within the first five iterations, before leveling off after approximately ten iterations. In particular, the predicted BBB improved for the first five iterations, while the predicted solubility began to level off after approximately ten iterations. This indicated that the compounds evolved exhibited a better balance of predicted properties. One example of a compound generated in the eighth iteration was presented and is shown in Figure 11(b). In [50] a further, more complex, example application was also presented corresponding to the simultaneous optimization of this compound to improve BBB, solubility and binding to Cytochrome P450 CYP2D6.

The authors suggest that tools like Pareto Ligand Designer and other methods for multi-parameter *de novo* design [40] [55] have the potential to generate relevant, synthesizable molecules while considering many properties simultaneously as a source for ideas in lead discovery, lead optimization and beyond.



Rat IC₅₀ = 0.30 μM

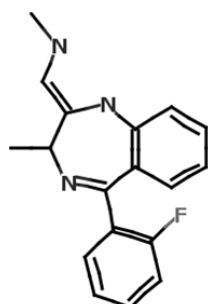
MW = 425.41

AlogP = 3.44

Predicted BBB (log([Brain]:[Blood])) = -0.58

Predicted log solubility (M) = -5.77

(a)



Tanimoto similarity to (a) = 0.35

MW = 295.35

AlogP = 2.875

Predicted BBB (log([Brain]:[Blood])) = 0.15

Predicted log solubility (M) = -4.22

(b)

Figure 11 Example compound structures published by Ekings *et al.* [50] as an illustration of the application of Pareto Ligand Designer. The structure in (a) is a known CCK antagonist reported by Evans *et al.* [51] which was predicted to have issues with respect to blood-brain barrier penetration and solubility. The structure in (b) is one of the compounds proposed by Pareto Ligand Designer that is predicted to have improved blood-brain barrier penetration and solubility while retaining sufficient similarity to the initial compound to suggest an acceptable probability of activity against CCK.

Rapid Focus in Lead Optimization using Probabilistic Scoring

Here we present an example of the application of the StarDrop software application [56] to an on-going drug discovery project with the objective of identifying an orally bioavailable compound for a CNS target. The original progress of the project, which did not use any of the MPO methods described herein, is outlined in Figure 12(a). This shows that the initial efforts were focused on a cluster of similar compounds in which good activity was identified, but the compounds exhibited either good CNS penetration or good oral bioavailability *in vivo*, but not both simultaneously.

As shown in Figure 12(b), scoring the 200 compounds that were initially progressed, against the profile shown in Figure 13(a) and using a probabilistic scoring algorithm, suggests that the chance of success of these compounds is very small, in agreement with the experimental results. Furthermore, analysis of the resulting scores suggests that an alternative region of the 'chemical space' would be more likely to yield compounds with an appropriate balance of properties. This was again supported by the practical experience of the project; the next set of 200 compounds that had been progressed was focused on this alternative region and identified highly potent compounds with an improved balance of CNS penetration and oral bioavailability, as shown in Figure 12(c). However, in order to achieve this approximately 3100 compounds had been synthesized and tested for *in vitro* potency, approximately 400 compounds were progressed for detailed *in vitro* ADME studies and approximately 70 compounds were studied with *in vivo* pharmacokinetic (PK) models. Furthermore, there were no obvious strategies to find new active compounds with additional improvements in the PK.

An alternative, MPO approach to this project was explored by retrospective application to the full library of 3,100 compounds explored by this project with the goal of selecting 25 compounds for *in vivo* study. An outline of the process is shown in Figure 14. Predictions for key ADME and physicochemical properties were made for the full library of 3,100 compounds and the compounds were scored using a probabilistic scoring algorithm against the profile shown in Figure 13(a) for a good balance of properties for an orally dosed compound against a CNS target. 300 compounds were selected using a genetic algorithm and an objective function corresponding to a weighted sum of score and diversity ($w_s=0.25$, $w_d=0.75$). In this case, more weight was assigned to diversity because little was known about the structure-activity relationship for target potency and therefore it was necessary to explore the full chemical space to maximize the chance of identifying diverse, potent compounds. The potency data for the 300 selected compounds then were used, along with the predicted properties to rescore these compounds for a balance of potency and appropriate ADME and physicochemical properties using the profile shown in Figure 13(b). Finally, based on these scores, 25 compounds were selected, again using a genetic algorithm to select based on a balance of score and diversity. However, in this case more weight was applied to score than diversity ($w_s=0.75$, $w_d=0.25$) to provide greater focus on the compounds with the highest chance of success.

The results of this process are summarized in Figure 12(d). This shows that the initial selection of 300 compounds provided a good coverage of the chemical space explored by this project and high scoring compounds were identified in different regions. The MPO process selected a number of compounds for which *in vivo* PK data had previously been determined and it is notable that the same results were found for the region that was heavily explored in the first phase of the project; good oral bioavailability or good CNS penetration, but not both. Furthermore, the best compound previously identified, with a good balance of potency, oral bioavailability and CNS penetration was also selected by MPO. Finally, a new region of chemical space was highlighted that had not previously been studied using *in vitro* ADME assays or *in vivo* PK, providing a new avenue for exploration for the project.

This example shows how an MPO approach can significantly improve the efficiency with which high quality compounds can be identified, by focusing early on the chemistries most likely to have a good balance of properties required for *in vivo* efficacy. In this case, the same information could have been derived through synthesis and *in vitro* testing of 90% fewer compounds and with 70% fewer *in vivo* PK studies than the traditional process. Furthermore, a broader range of strategies were explored resulting in more opportunities to identify a good candidate drug.

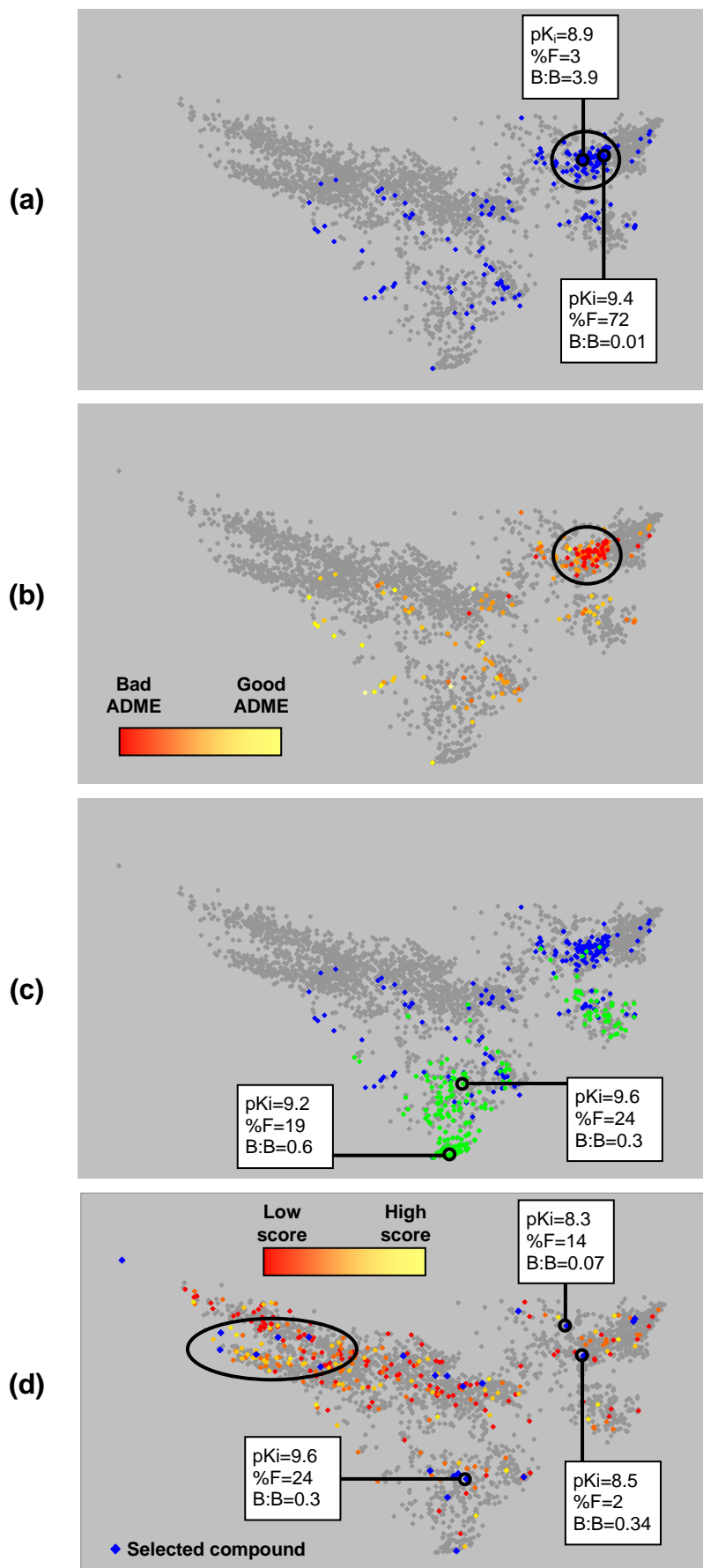


Figure 12 Chemical space plots (as defined in Figure 9) illustrating the selection of compounds from a project targeting an orally active compound for a CNS target, as described in the text. Plots (a) through (c) show the initial progress of the project. In plot (a) the first 200 compounds selected for progression for detailed *in vitro* ADME testing are highlighted in blue. These are highly focused on a small region of chemical space shown by an ellipse and illustrative examples of the properties for two compounds are shown for *in vitro* activity against the target (pK_i), and *in vivo* oral bioavailability (%F) and compound concentration ratio between brain and blood (B:B). Plot (b) shows the probabilistic scores for these compounds scored for an appropriate profile of predicted ADME properties for an orally dosed compound for a CNS target shown in Figure 13(a); the points are colored from low (red) to high (yellow) according to their ADME scores, indicating that the region shown in the ellipse corresponds to high risk chemistry, while other regions are more likely to yield a good balance of properties. Plot (c) shows the second 200 compounds chosen by the project for progression in green and the experimental results shown for two example compounds confirm that this region yielded compounds with improved *in vivo* disposition, as predicted. Finally, plot (d) illustrates an alternative strategy for exploration of this chemical space, as outlined in Figure 13. The compounds selected for potency testing, based on a balance of score and diversity, are colored from red to yellow, corresponding to low to high score respectively for a balance of *in vitro* potency and predicted ADME properties using the profile in Figure 13(b). The dark blue points indicate 25 compounds selected for *in vivo* PK testing and results for three illustrative compounds are shown. This suggested an alternative area of chemical space, shown by the ellipse, which had not previously been investigated with *in vitro* ADME and *in vivo* PK.

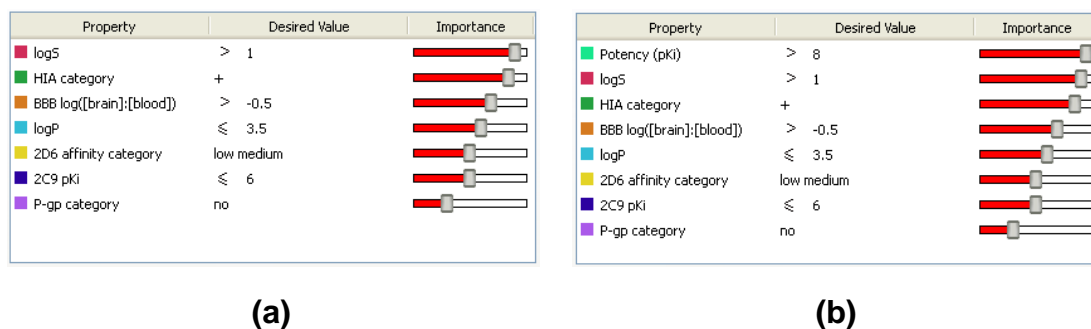


Figure 13. The profiles used to score compounds for an oral administration against a CNS target in the process shown in Figure 14. Profile (a) employs only predicted ADME properties that may be applied to a virtual library before target activity has been measured. The profile shown in (b) combines the predicted ADME profile with the experimentally measured target potency (pKi) to prioritize compounds that have been synthesized and tested for target activity for further study.

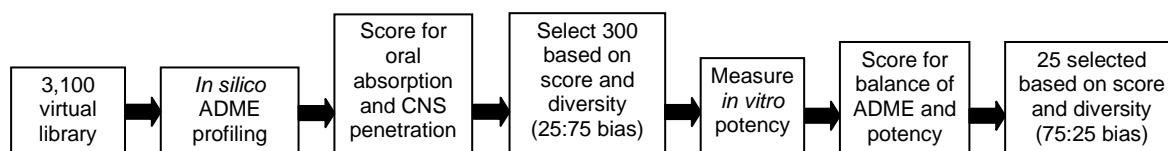


Figure 14. A process for exploration of the project chemical space illustrated in Figure 12, based on probabilistic scoring and compound selections that balance quality and diversity, as described in detail in the text.

Conclusions

It is well established that successful drugs require a delicate balance of many factors relating to their biological and physicochemical properties and achieving an appropriate combination of properties is one of the major challenges in the drug discovery process. Dealing effectively with complex, multi-dimensional data is a challenge. Therefore, there is a natural tendency to simplify and initially focus on optimization of a critical property, typically target potency. The expectation is that, having optimized this property, other properties can be tackled either by filtering compounds that do not meet the corresponding criteria or optimizing these in turn. Unfortunately, having become locked into a tight area of chemistry by the structure-activity relationship, it is often impossible to optimize other properties without sacrificing potency. This leads to late stage failures or multiple, long iterations in lead optimization due to the need to ‘lead hop’ to new chemical series in the search for acceptable *in vivo* disposition and safety. To avoid these issues, it is important to simultaneously consider multiple parameters as early as possible in the process to quickly focus on chemistries with a good balance of properties that give the best chance for rapid progress in lead optimization.


In this review, we have explored a range of approaches to guide the design and selection of compounds that achieve a required multi-parameter profile, from simple rules-of-thumb through to more sophisticated and computational methods. We have examined the relative strengths and weaknesses of these approaches relative to the requirements for interpretability, flexibility with respect to setting property criteria and their weights, and the capability to take into account the uncertainty in the underlying data on which decisions are based. A summary of these relative strengths and weaknesses is summarized in Table 3.

Table 3. A summary of the six MPO methods discussed and relative strengths and weaknesses against the requirements for application to drug discovery.

	Interpretability	Flexibility in Criteria	Weighting	Account for Uncertainty
Rules of Thumb	Very easy to understand	Criteria are pre-determined	No weighting	Only by applying criteria with discretion
Filtering	Very easy to understand	Cut-off criteria may be arbitrarily defined	Only by specifying order of filters	No
Calculated Metrics	Clear interpretation except for most complex metrics	Metric is pre-defined. Flexibility in setting target value	Pre-defined by metric	Uncertainty rarely propagated to value of calculated metric
Pareto Optimization	Clear interpretation for small number of properties	Not based on criteria. Different property balances selected	Difficult to weight different properties in algorithm. Post-hoc analysis can be applied	Not explicitly considered, but spreads risk across different strategies
Desirability Functions	Impact of each property easy to interpret	Arbitrary criteria may be defined. Not only hard cut-offs	Arbitrary weights may be applied to each property criterion	Avoiding hard cut-offs reduces impact of uncertainty
Probabilistic Scoring	Impact of each property easy to interpret	Arbitrary criteria may be defined. Not only hard cut-offs	Arbitrary weights may be applied to each property criterion	Explicitly accounts for uncertainty and demonstrates its impact

Choosing the most appropriate MPO method for a drug discovery project depends the project's phase and objectives and the availability of data. When little experimental or predicted property data is available, for example early in a project, and the goal is to identify a typical orally bioavailable, small molecule drug, it makes sense to learn from the history of successes and failures in drug discovery by applying appropriate rules of thumb or calculated metrics. These will help to bias the odds in favor of success by guiding the chemistry toward an appropriate physicochemical property space. If the property requirements for a successful compound are not known *a priori*, Pareto optimization provides a powerful approach to explore different trade-offs in the search for the best balance of properties. This method is particularly useful in designing libraries or expanding chemistry to explore a wide range of different strategies. In this scenario, considering the uncertainty in the data is less important, as the objective is to 'spread the bets' across a wide range of possibilities. Indeed, when the downstream results following a Pareto optimization indicate the most appropriate property profile, a more focused approach can be used to reexamine the options and identify alternatives that are likely to satisfy these requirements. Finally, when a project team is able to define a target product profile for an ideal compound and appropriate trade-offs between the property criteria, a method based on desirability functions will help to quickly focus on the compounds with the best balance of properties, without the risk of the artificially hard distinctions between compounds that filters would impose. Ideally, in this scenario, a probabilistic approach should be applied to avoid giving undue weight to uncertain data and to highlight where obtaining additional data would permit a more confident decision to be made.

The MPO methods discussed in this review enable predicted and/or experimental data to be integrated and assessed against a project's objectives to prioritize compounds throughout the drug discovery process. We have presented some illustrative examples of how MPO can be applied from hit discovery through lead optimization to identify compounds with the best chance of success against a project's objectives. Most of the examples presented in this review have focused on the application of MPO to properties data generated *in*



silico. This is probably due to the origins of MPO as a computational field, hence early adopters in the drug discovery community have tended to come from this background. In addition, the strongest need for MPO arises when interpreting large quantities of data and *in silico* methods have the greatest potential to generate data sets containing large numbers of compounds and properties, quickly and at low cost. However, it should be emphasized that most of the MPO methods discussed herein can equally be applied to purely experimental data. This is becoming increasingly important given the range of assays for ADME and toxicity properties that are routinely conducted on large numbers of compounds in early drug discovery. An example application of probabilistic scoring to an data set comprised only of *in vitro* data, to identify compounds with improved *in vivo* disposition in lead optimization, is described in [33].


Finally, given the complexity of the data that is considered in MPO, most of these methods rely on computational algorithms and software. Simply gathering, storing and making the data easily available to project scientists is a challenge for informatics platforms [57]. However, it is essential that software implementing MPO algorithms be intuitive and user friendly, as it should be possible for all decision makers to explore trade-offs in data and easily interpret the results, even if they are not computational experts. These tools should also provide a foundation for the effective collaboration of scientists from the different disciplines that contribute to drug discovery, as each will bring a different perspective to the properties required for a successful compound.

Acknowledgements


The author would like to thank current and former colleagues for valuable discussions and critical contributions to some of the methods described herein. These include, but are not limited to, Ed Champness, Chris Leeding, Iskander Yusof, James Chisholm, Mike Tarbit, Alan Beresford, Dawn Yates, Brett Saunders, Dan Hawksley and Olga Obrezanova. I would also like to thank the organizers and participants in the “Optimizing Drug Design” workshop held at the Lorentz Center of the University of Leiden, Netherlands in July 2009, at which many of these methods and concepts were discussed. The proceedings of this workshop can be found at <http://www.lorentzcenter.nl/lc/web/2009/359/info.php3?wsid=359>.

References

- 1 Paul S, Mytelka D, Dunwiddie D, Persinger C, Munos B, Lindborg S, Schacht A. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* 2010;9:203-14.
- 2 Chadwick AT, Segall MD. Overcoming psychological barriers to good discovery decisions. *Drug Discov. Today.* 2010;15:561-569.
- 3 Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates. *Nature Rev. Drug Discov.* 2004;3:711-716.
- 4 Dimasi J, Hansen R, Grabowski H. The price of innovation: new estimates of drug development costs. *J. Health Econ.* 2003;22:151-85.
- 5 Barnett Educational Services. PAREXEL Biopharmaceutical R&D Statistical Sourcebook 2010/2011. Waltham: Parexel International Corporation; 2010.
- 6 Segall MD. Why is it still Drug Discovery? *European Biopharmaceutical Review.* 2008.
- 7 Segall MD, Champness E. The difference between guiding and supporting decisions: Enhancing decisions and improving success in drug discovery. *Genetic Engineering News.* 2010 September.
- 8 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.
- 9 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;45:2615-2623.
- 10 Lu JJ, Crimin K, Goodwin JT, Crivori P, Orrenius C, Xing L, Tandler PJIVTJ, Amore BM, Wilson AGE, Stouten PFW, et al. Influence of Molecular Flexibility and Polar Surface Area Metrics on Oral Bioavailability in the Rat. *J. Med. Chem.* 2004;47:6104-6107.
- 11 Johnson TW, Dress KR, Edwards M. Using the Golden Triangle to optimize clearance and oral absorption. *Bioorg. Med. Chem. Lett.* 2009;19:5560-5564.
- 12 Lovering F, Bikker J, Humblet C. Escape from flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* 2009;52:6752-6756.
- 13 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;14:1011-1020.
- 14 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;18:4872-4875.
- 15 Choy YB, Prausnitz MR. The rule of five for non-oral routes of drug delivery: ophthalmic, inhalation and transdermal. *Pharm. Res.* 2011;28:943-948.
- 16 Hidalgo IJ, Raub TJ, Borchardt RT. Characterization of the human colon carcinoma cell line (Caco-2) as a model system for intestinal epithelial permeability. *Gastroenterology.* 1989;96:736-749.
- 17 Schoonen WG, de Roos JA, Westerink WM, Débiton E. Cytotoxic effects of 110 reference compounds on HepG2 cells and for 60 compounds on HeLa, ECC-1 and CHO cells. II mechanistic assays on NAD(P)H, ATP and DNA contents. *Toxicol. In Vitro.* 2005;19:491-503.

- 
- 18 Hopkins AL, Groom CR, Alexander A. Ligand Efficiency: a useful metric for lead selection. *Drug Discov. Today.* 2004;9:430-431.
 - 19 Carr RAE, Congreve M, Murray CW, Rees DC. Fragment-based lead discovery: leads by design. *Drug Discov. Today.* 2005;10:987-992.
 - 20 Leeson PD, Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.* 2007;6:881-890.
 - 21 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;18:4872-4875.
 - 22 Edwards MP, Price D. Role of physicochemical properties and ligand lipophilicity efficiency in addressing drug safety risks. *Annual Reports in Medicinal Chemistry.* 2010;45:381-391.
 - 23 Abad-Zapatero C. Ligand efficiency indices for effective drug discovery. *Expert Opin. Drug Discov.* 2007;2:469-488.
 - 24 Keseru GM, Makara GM. The influence of lead discovery strategies on the properties of drug candidates. *Nat. Rev. Drug Discov.* 2009;8:203-212.
 - 25 Reynolds CH, Bembenek SD, Tounge BA. The role of molecular size in ligand efficiency. *Bioorg. Med. Chem. Lett.* 2007;17:4285-4261.
 - 26 Ladbury JE, Klebe G, Freire E. Adding calorimetric data to decision making in lead discovery: a hot tip. *Nature Rev. Drug Discov.* 2010;9:23-27.
 - 27 Ferenczy GG, Keseru GM. Thermodynamics guided lead discovery and optimization. *Drug Discov. Today.* 2010;15:919-932.
 - 28 Jaffe W. Pareto translated: A review article. *J. Economic Literature.* 1972;10:1190-1201.
 - 29 Kruisselbrink J, Michael E, Back T, Bender A, Ijzerman A, van der Horst E, Ehrgott M, Fonesca C, Gandibleux X, Hao JK, et al. Combining aggregation with Pareto optimization: A case study in evolutionary molecular design. In: *Proceedings of EMO '09: Lect. Notes Comput. Sci.*; 2009; Berlin/Heidelberg. p. 453-467.
 - 30 Harrington EC. The desirability function. *Ind. Qual. Control.* 1965;21:494-498.
 - 31 Derringer G, Suich R. Simultaneous optimization of several response variables. *Journal of Quality Technology.* 1980;12:214-219.
 - 32 Segall MD, Champness E, Obrezanova O, C. L. Guiding the decision-making process to identify high quality compounds. *Drug Metab. Rev.* 2009;41:7-186 Abstract 244.
 - 33 Segall M, Beresford A, Gola J, Hawksley D, MH T. Focus on success: using a probabilistic approach to achieve an optimal balance of properties in drug discovery. *Expert Opin. Drug Metab. Toxicol.* 2006;2:325-37.
 - 34 Segall M, Champness E, Obrezanova O, Leeding C. Beyond profiling: Using ADMET models to guide decisions. *Chemistry & Biodiversity.* 2009;6:2144-2151.
 - 35 Hopkins AL. *The Practice of Medicinal Chemistry.* London: Academic Press; 2008. p. 521-532.

- 36 Bohacek RS, McMartin C, Guida WC. The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.* 1996;16:3-50.
- 37 Clark DE, editor. *Evolutionary Algorithms in Computer-Aided Molecular Design*. Weinheim: Wiley-VCH; 2000.
- 38 Brown N, McKay B, Gilardoni F, Gasteiger J. A graph-based genetic algorithm and its application to the multiobjective evolution of median molecules. *J. Chem. Inf. Comput. Sci.* 2004;44:1079-1087.
- 39 Stewart K, Shiroda M, James C. Drug Guru: a computer software program for drug design using medicinal chemistry rules. *Bioorg. Med. Chem.* 2006;14:7011-22.
- 40 Segall MD, Champness EJ, Leeding C, Lilien R, Mettu R, Stevens B. Applying medicinal chemistry transformations to guide the search for high quality leads and candidates. *J. Chem. Inf. Model.* 2011;(in press).
- 41 Brown N, McKay B, Gasteiger J. A novel workflow for the inverse QSAR problem using multiobjective optimization. *J. Comp. Aided Mol. Design.* 2006;20:333-341.
- 42 Boda K, Seidel T, Gasteiger J. Structure and reaction based evaluation of synthetic accessibility. *J. Comp. Aided Mol. Des.* 2007;21:311-325.
- 43 Nicolaou CA, Brown NA, Pattichis CS. Molecular optimization using computational multi-objective methods. *Curr. Opin. Drug Discov. Devel.* 2007;10:316-324.
- 44 Rogers DJ, Tanimoto TT. A computer program for classifying plants. *Science.* 1960;132:1115-1118.
- 45 Agrafiotis DK. Diversity of chemical libraries. In: Schleyer P, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer III HF, P.R. S, editors. *The Encyclopedia of Computational Chemistry*. Vol 1. Chichester: John Wiley and Sons; 1998. p. 742-761.
- 46 Jones G. Genetic and evolutionary algorithms. In: Schleyer P, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer III HF, P.R. S, editors. *Encyclopedia of Computational Chemistry*. Chichester: John Wiley and Sons; 1998.
- 47 Gillet VJ, Khatib W, Willett P, Fleming PJ, Green DVS. Combinatorial library design Using a multiobjective genetic algorithm. *J. Chem. Inf. Comput. Sci.* 2002;42:375-385.
- 48 Fonesca CM, Fleming PJ. Genetic algorithms for multiobjective optimization: formulation, discussion and generalisation. In: Forrest S, editor. *Genetic Algorithms: Proceedings of the Fifth International Conference*; 1993; San Mateo, CA. p. 416-423.
- 49 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;1:435-449.
- 50 Ekins S, Honeycutt JD, Metz JT. Evolving molecules using multi-objective optimization: applying to ADME/Tox. *Drug Discov. Today.* 2010;15:451-460.
- 51 Evans BK, Rittle KE, Bock MG, DiPardo RM, FRM, WWL., Lundell GF, Veber DF, Anderson PS. Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. *J. Med. Chem.* 1988;31:2235-2246.
- 52 Rogers D, Brown RD, Hahn M. Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis in high-throughput screening follow-up. *J. Biomol. Screen.* 2005;10:682-686.

- 
- 53 Muchmore SW, Debe DA, Metz JT, Brown SP, Martin YC, Hajduk PJ. Application of belief theory to similarity data fusion for use in analog searching and lead hopping. *J. Chem. Inf. Model.* 2008;48:941-948.
- 54 Metz JT, Huth JR, Hajduk PJ. Enhancement of chemical rules for predicting compound reactivity towards protein thiol groups. *J. Comput. Aided Mol. Des.* 2007;21:139-144.
- 55 Nicolaou CA, Apostolakis J, Pattichis CS. De novo drug design using multiobjective evolutionary graphs. *J. Chem. Inf. Model.* 2009;49:295-307.
- 56 Optibrium. [Internet]. [cited 2011 March 3]. Available from: <http://www.optibrium.com/stardrop>.
- 57 Lusher SJ, McGuire R, Azevedo R, Boiten JW, van Schaik RC, de Vlieg J. A molecular informatics view on best practice in multi-parameter compound optimization. *Drug Discov. Today.* 2011;16:555-568.