

Rules for drug discovery: can simple property criteria help you to find a drug?

The years since the publication of Lipinski's Rule of Five (Ro5)¹ in 1997 have seen the growth of a minor industry, dedicated to generating new 'rules' for drug discovery. Fuelled by what Kenny and Montanari have termed "Ro5 envy"², these rules are usually based on simple calculated compound characteristics and define criteria for the selection of compounds in drug discovery. But, can a generic rule be used to reliably select compounds for drug discovery?

'Rules of thumb' for selection of 'drug-like' compounds have been defined for a wide range of properties, including: lipophilicity (logP and logD), molecular weight (MW), number of hydrogen bond donors and acceptors (HBD and HBA), number of rotatable bonds (ROTB), polar surface area (PSA), fraction of sp³ carbons (FSP3), number of aromatic rings (AROM), acid dissociation coefficient (pKa) and counts of alerts for undesirable functional groups (ALERT)³. Criteria for these properties have been related to a diverse range of outcomes including oral bioavailability^{1,4}, safety⁵, 'developability'⁶ and clinical success⁷.

One common misconception is that these rules define general property criteria that describe a 'drug like' compound. However, it is important to remember that, in most cases, they have been developed with a particular objective in mind; in the case

of the Ro5, the property criteria (logP < 5, MW < 500, HBD < 5, HBA < 10) were proposed as a rule for selection of orally administered drugs on the basis of properties related to solubility and permeability. Application to projects with other intended routes of administration may be misleading.

The simplicity of these rules and the ease with which they may be applied make them attractive. However, this very simplicity and memorability may give rise to unconscious bias in decisions made in compound optimisation. In a recent paper⁸, Michael Shultz coined the term "Lipinski's Anchor", referring to a psychological effect known as 'anchoring', whereby a number suggested before making an estimate of an unknown value exerts a significant influence on the outcome of the estimate. He notes that the simple cut-off values proposed by these rules may serve as anchors that have a disproportionate effect on

By Dr Matthew Segall

References

- 1 Lipinski, CA, Lombardo, F, Dominy, BW and Feeney, PJ (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23, 3-25.
- 2 Kenny, PW and Montanari, CA (2013). Inflation of correlation in the pursuit of drug-likeness. *J. Comput.-Aided Mol. Des.* 27, 1-13.
- 3 Garcia-Sosa, AT, Maran, U and Hetenyi, C (2012). Molecular Property Filters Describing Pharmacokinetics and Drug Binding. *Curr. Med. Chem.* 19, 1646-1662.
- 4 Veber, DF, Johnson, SR, Cheng, HY, Smith, BR, Ward, KW and Kopple, KD (2002). Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* 45, 2615-2623.
- 5 Hughes, JD, Blagg, J, Price, DA et al (2008). Physicochemical drug properties associated with in vivo toxicological outcomes. *Bioorg. Med. Chem. Lett.* 18, 4872-4875.
- 6 Ritchie, TJ and Macdonald, SJF (2009). The impact of aromatic ring count on compound developability – are too many aromatic rings a liability in drug design? *Drug Discov. Today* 14, 1011-1020.
- 7 Lovering, F, Bikker, J and Humblet, C (2009). Escape from flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* 52, 6752-6756.
- 8 Shultz, MD (2013). Improving the Plausibility of Success with Inefficient Metrics. *ACS Med. Chem. Lett.* 5, 2-5.
- 9 Bickerton, GR, Paolini, GV, Besnard, J, Muresan, S and Hopkins, AL (2012). Quantifying the chemical beauty of drugs. *Nature Chemistry* 4, 90-98.

Table 1: Summary of the performance of the Ro5¹ and rules derived by Veber et al (ROTB < 10 and PSA < 140 Å²)⁴ when applied to distinguish drugs approved for oral administration from those administered by non-oral routes

ROUTE OF ADMINISTRATION	RO5 ¹		VEBER ET AL ⁴	
	Pass	Fail	Pass	Fail
Oral	709	59	639	130
Non-oral	333	90	267	156

decisions in drug discovery projects. Shultz further notes that: “When judging probability, people rely on representativeness heuristics (a description that sounds highly plausible), while base-rate frequency is often ignored.”

The correlations between the simple properties on which these rules are based and the *in vivo* disposition of a compound are typically weak. Therefore, the predictive performances of these rules are limited, suggesting that they should be

used with caution. For example, Table 1 shows the performance of two commonly applied rules for selecting orally bioavailable drugs to a data set of 1,191 approved drugs, labelled as oral or non-oral depending on their approved routes of administration. One should be careful not to over interpret the results from such a biased set; however, we can see that passing either rule is not a guarantee of finding an orally bioavailable compound, more non-orally administered compounds pass the rules

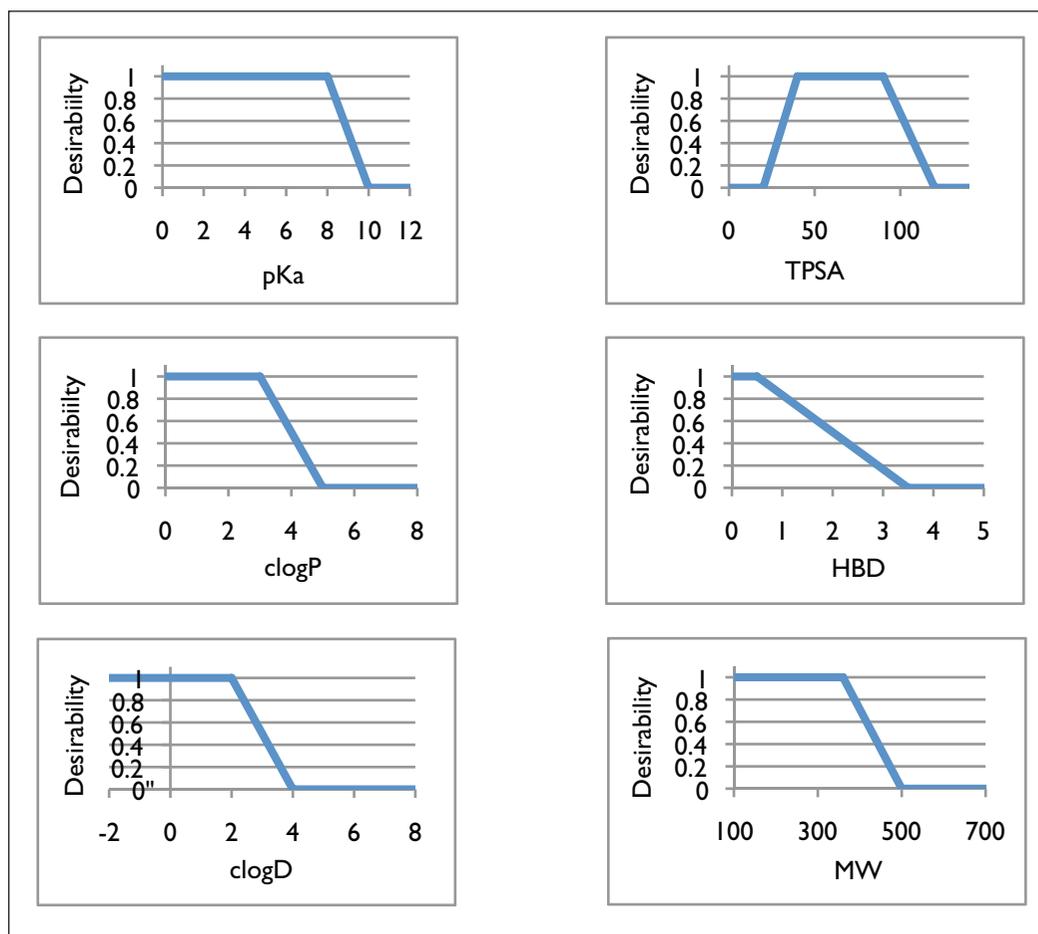


Figure 1: Six desirability functions used for the calculation of the CNS MPO score¹⁰. Each relates a property value to its 'desirability' from ideal (1.0) to completely unacceptable (0.0)

Continued on page 24

than fail and a significant proportion of compounds that fail the Ro5 are orally administered.

Hard cut-offs draw artificially harsh distinctions between compounds with similar properties; does a compound with a logP of 4.9 really have a significantly lower chance of success than one with a logP of 5.1? This is further exacerbated by the fact that some of these parameters have significant uncertainty, for example predictions of logP typically have an uncertainty of approximately ± 0.5 log units, making the difference between the values 4.9 and 5.1 statistically insignificant.

'Drug likeness' metrics

In order to avoid the problems of hard cut-offs, continuous metrics have been proposed that generate a numerical value on the basis of which compounds can be ranked. These typically use 'desirability functions' that map the value of each property onto a scale between 0 and 1 that represents the 'desirability' of that outcome (some simple examples are shown in Figure 1). An ideal value of a property will achieve a desirability of 1 and the worst possible values have a desirability of 0. The desirabilities of multiple properties are then added or multiplied to generate a single value representing the quality of the compound. In some cases arithmetic or geometric means are used to normalise for the number of properties.

A popular example of this approach is the Quantitative Estimate of Drug-likeness (QED) that combines desirability functions for eight properties, MW, clogP, HBD, HBA, PSA, ROTB, AROM and ALERT⁹. The desirability function for each property was fitted to the distribution of the property for 771 oral drugs, such that the highest desirability is given to the property values corresponding to the highest proportion of oral drugs. The overall QED is calculated by taking the geometric mean of the resulting eight desirabilities (the eighth root of their product) to give a number between 0 and 1, representing the similarity of the properties of a compound to those of the majority of oral drugs.

A similar approach was proposed by Wager et al¹⁰ for the selection of compounds with an improved chance of success as a drug intended for a target in the central nervous system (CNS). In this case, desirability functions were constructed for six properties (MW, logP, logD, PSA, HBD and pKa of the most basic nitrogen) based on the experience of project scientists working on discovery of CNS drugs (Figure 1). The 'CNS MPO score' is calculated by adding the desirabilities of the individual properties to give a number between 0 and 6.

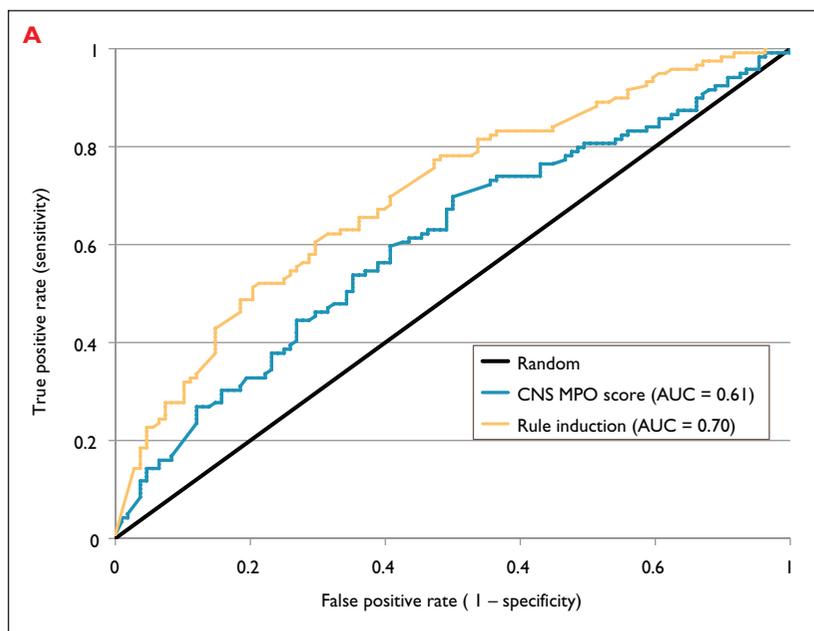
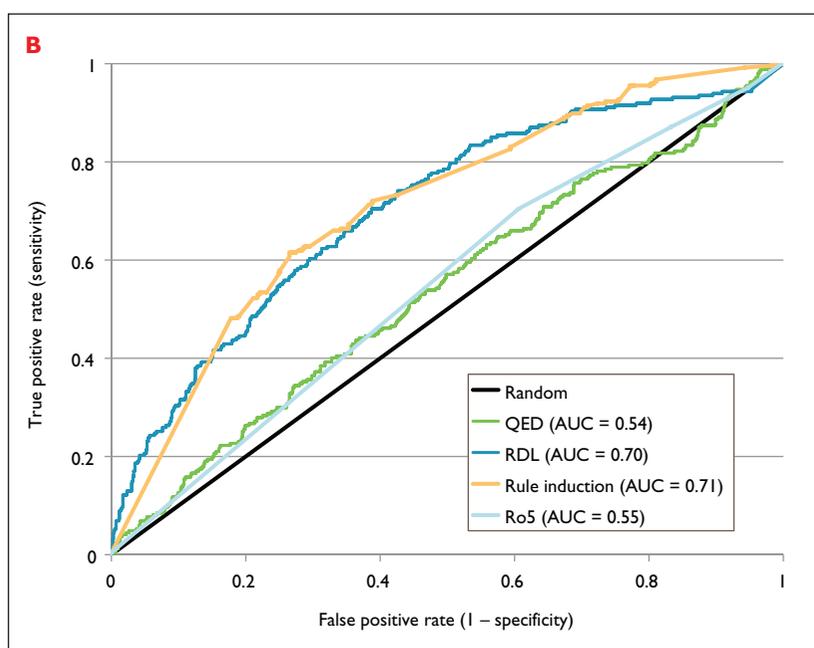


Figure 2: ROC plots of the true positive rate (TPR (sensitivity)) against the false positive rate (FPR (1 – specificity)) for the classification of compounds. A perfect classifier would be represented by the point in the top left and a performance below the identity line indicates worse performance than a random classification. A greater area under the curve (AUC) for a classifier indicates higher performance, a random selection will have an AUC of 0.5 and the maximum AUC is 1.

A ROC curves for selection of CNS drugs from a data set containing 119 marketed CNS drugs and 108 failed Pfizer candidates derived from reference ¹⁰. Curves are shown for the CNS MPO score and a profile generated by rule induction and shown in Figure 4(a).

B ROC curves for classification of compounds as orally absorbed drugs or otherwise using RDL, QED, and a profile generated by rule induction and shown in Figure 4(b). The curve corresponding to the Ro5 is also shown for comparison. In this case, a set of 247 orally administered drugs was differentiated from 1,000 randomly selected compounds from ChEMBL



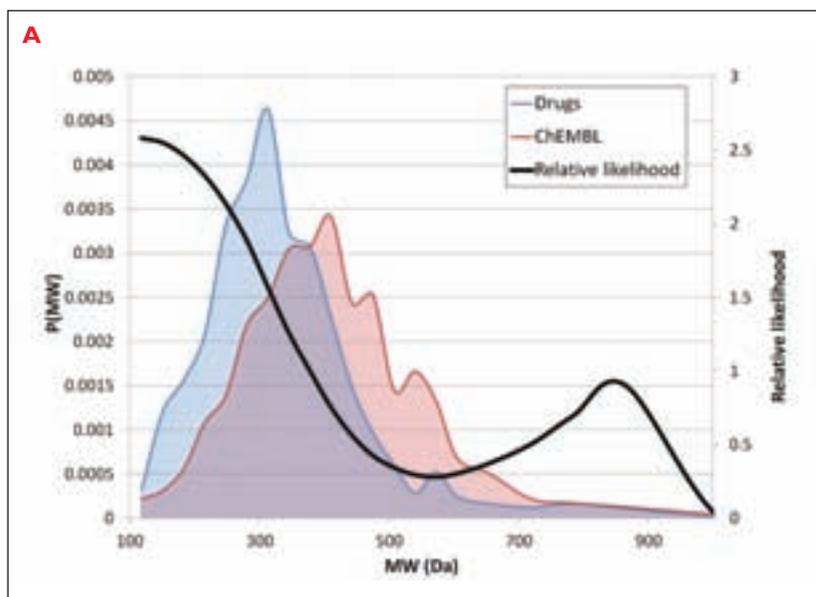
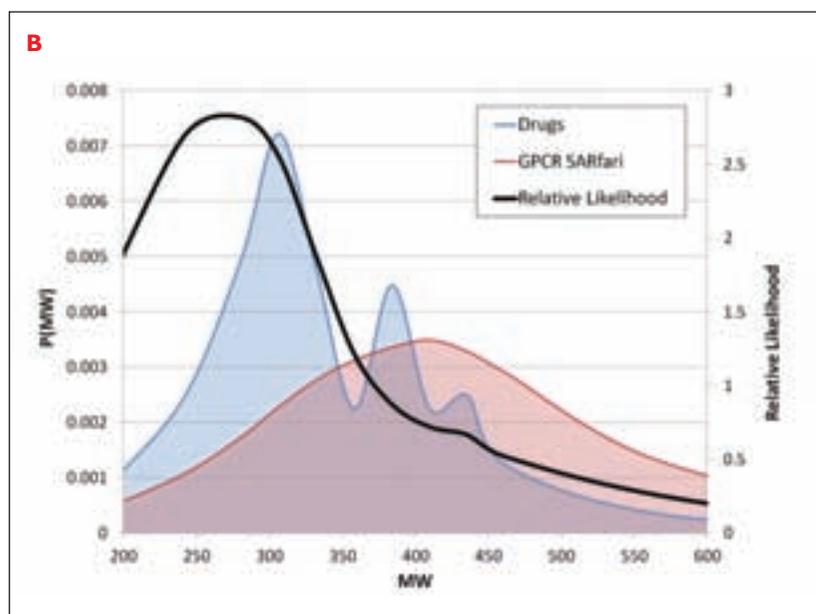


Figure 3: Graphs showing distributions of MW for oral drugs (blue) and non-drugs (red). The resulting 'likelihood function', indicating the relative likelihood of a compound being an oral drug (right vertical axis), is shown as a black line.

A Shows the distributions and resulting likelihood function for a set of 771 oral drugs compared with 1,000 randomly selected compounds from the ChEMBL database.

B Shows the distributions and resulting likelihood function for orally administered drugs for GPCR targets (blue) and compounds screened against GPCR targets in the GPCR SARfari database (red)



In both of these cases the approach seems plausible; it intuitively makes sense that a compound that is similar to known oral drugs would be more likely to be a successful oral drug and that experienced scientists will understand the properties that improve the chance of finding a CNS drug. However, we should recall Shultz's observation that plausibility should not be substituted for an understanding of the underlying statistics.

In the case of the CNS MPO score, the authors calculated the scores for compounds in a data set of 119 marketed CNS drugs and 108 failed Pfizer candidates for CNS indications and found that 74% of the drugs had a CNS MPO > 4, but only 60% of the failed candidates. While this sounds promising, it should be noted for comparison that, in the same data set, 75% of drugs have a MW < 350 while only 44% of failed candidates meet this threshold; a better discrimination than provided by the rule CNS MPO > 4. We would not suggest selecting potential CNS drugs based on a simple MW cut-off!

A rigorous approach for comparing the performance of a metric for the selection of compounds is provided by receiver operating characteristic (ROC) plots, as shown in Figure 2. The result of applying the CNS MPO score to the data set of CNS drugs and failed candidates published by Wager et al is shown in Figure 2(a), showing that the performance on this set is not much better than random selection. We can similarly investigate the performance of QED for discrimination of 247 oral drugs (different from those used to fit the desirability functions) from 1,000 randomly selected non-drug compounds from the ChEMBL database of compounds published in medicinal chemistry journals¹¹. The resulting ROC curve is shown in Figure 2(b), showing that the performance in this case does not differ significantly from random. Similar results were found by Debe et al when they applied the CNS MPO score and QED in an attempt to discriminate between 250 marketed neuroscience drugs and a background set of 'leads' (compounds with micromolar or better inhibition of a drug target)¹².

Similarity or difference?

Here we come back again to the point raised by Shultz; it is necessary to compare the property distributions of successful compounds with the base-rate frequency. In other words, we are interested in the properties that make a successful compound *different* from other compounds, not simply in the properties that successful compounds have in common.

To illustrate this, Yusof et al developed a measure called the Relative Drug Likelihood (RDL) to identify property values that increase the likelihood of finding a successful compound, by comparing the properties of successful compounds with those of a ‘background’ set of unsuccessful compounds¹³. While this approach can be applied to any objective and properties, for the purposes of comparison, the authors explored the same six properties and set of 771 oral drugs as used for the QED analysis, and compared these with 1,000 non-drugs randomly selected from the ChEMBL database. An example of the resulting ‘likelihood function’ for MW is shown in Figure 3(a). The likelihoods for the six properties were combined by taking their geometric mean to calculate an overall RDL value. This was then applied to distinguish an independent set of 247 oral drugs from a different set of 1,000 randomly selected non-drugs from ChEMBL. The resulting ROC curve is also shown in Figure 2(b), which shows a significantly better performance than QED for this challenge.

The authors of the RDL paper also investigated the impact of considering drugs for a specific target class on the property values that give the highest likelihood of success. They considered oral drugs acting via G-protein coupled receptor (GPCR) targets and compared these with unsuccessful compounds screened against GPCR targets. Unsurprisingly, the results indicate that the property requirements for a GPCR drug differ significantly from the generic property requirements for an oral drug when considered across multiple target classes and therapeutic indications. This finding is likely to be repeated if we explore different therapeutic indications, routes of administration or target classes and casts further doubt on the value of a generic rule for the selection of ‘drug-like’ compounds.

Tailored multi-parameter rules

To address the need to find property rules tailored to specific project objectives, Yusof et al introduced a new method called ‘rule induction’ in a subsequent paper¹⁴. This helps to explore existing data to find multi-parametric rules that give the best improvement in the chance of success over the ‘background’ of unsuccessful compounds. The resulting rules are easy to interpret, allowing an expert to understand the property criteria and, if necessary, adjust them based on their understanding of the underlying biology and chemistry. The property requirements can be expressed as desirability functions, avoiding the drawbacks of hard cut-offs. Furthermore, the importance of each cri-

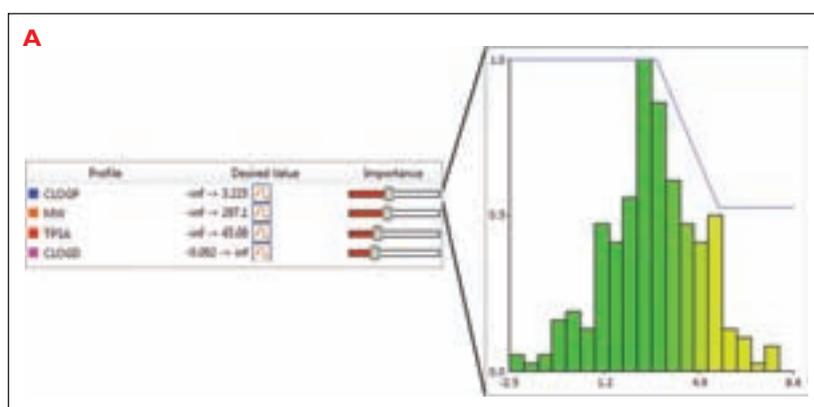
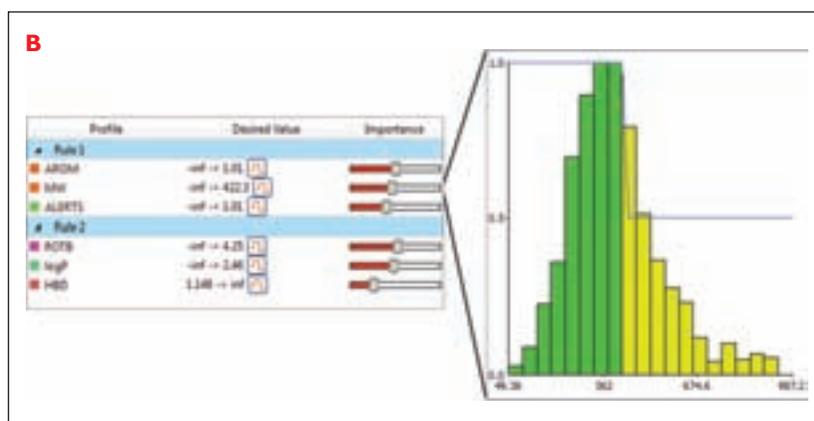


Figure 4: Profiles of property criteria derived using rule induction, showing the property name, ideal property values and importance for each criterion. Underlying each criterion is a desirability function, as illustrated for logP in **A** and MW in **B**.

A is a profile derived for discrimination of CNS drugs from failed candidates for CNS indications.

B is a profile for the selection of oral drugs. This profile is made up of two separate rules, each based on multiple property criteria. The score for a compound will be the highest derived from either of the rules



terion to the discrimination of successful and unsuccessful compounds can be found to indicate the most critical property requirements and potential trade-offs that can be made.

Examples of property rules from applying rule induction to the data sets used to define the CNS MPO and QED desirability functions are shown in Figure 4. It is notable that these rules only use a subset of the properties included in the CNS MPO and QED metrics respectively. Despite this, as shown in the ROC curves in Figure 3, the performances of the rules derived by rule induction are better than CNS MPO and QED, indicating that the omitted properties do not add any further value to the discrimination of successful and unsuccessful compounds. This highlights another issue with many published property rules; namely that they often include multiple properties that are

Continued from page 20

- 10** Wager, TT, Hour, X, Verhoest, PR and Villalobos, A (2010). Moving beyond rules: The development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. *ACS Chem. Neurosci.* 1, 435-449.
- 11** Gaulton, A, Bellis, LJ, Bento, AP et al (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40, D1100-D1107.
- 12** Debe, DA, Mamidipaka, RB, Gregg, RJ, Metz, JT, Gupta, RR and Muchmore, SW (2013). ALOHA: a novel probability fusion approach for scoring multi-parameter drug-likeness during the lead optimization stage of drug discovery. *J. Comput.-Aided Mol. Des.* 27, 771-82.
- 13** Yusof, I and Segall, MD (2013). Considering the impact drug-like properties have on the chance of success. *Drug Discovery Today* 18, 659-666.
- 14** Yusof, I, Shah, F, Hashimoto, T, Segall, MD and Greene, N (2014). Finding Rules for Successful Drug Optimization. *Drug Discov. Today*.

correlated. This commonly results from considering each property individually and combining the resulting criteria post-hoc to calculate an overall score, instead of explicitly considering the impact of the property criteria *in combination*. Including multiple, correlated properties in the calculation of a metric can lead to ‘overcounting’ of a single factor, artificially biasing the selection of compounds.

Conclusion

Simple property rules for selection and design of potential drugs may provide useful guidelines that help to avoid venturing into high risk property space, eg large, lipophilic compounds. However, in this article we have discussed a number of important caveats that should be taken into account, including:

- Hard cut-offs or filters draw artificially harsh distinctions between similar compounds and can lead to missed opportunities.
- The rules usually relate to a specific project objective and may not apply to your specific project objective.
- Metrics based on similarity with successful drugs do not take into account the property differences with the background of unsuccessful compounds.
- Inclusion of multiple, correlated properties can lead to over-counting of the same risk factor.

It would be ideal if there were a straightforward rule, based on simple properties, that could be used to identify compounds with a significantly higher chance of success as a drug. However, it seems clear that no single generic rule can fulfil this role. It is important to consider rules tailored to the specific objective of each project and tools are available that help to find and validate these rules.

It is also notable that none of the rules based on simple ‘drug-like’ properties strongly distinguish between successful drugs and other compounds, as illustrated by the ROC curves in **Figure 3**. The area under the curve (AUC) is a measure of the overall performance – an AUC of 0.5 is equivalent to random selection and an AUC of 1 signifies a perfect classifier – and the highest AUC for the methods investigated here is 0.7. This further emphasises that simple ‘drug-like’ properties do not have a strong correlation with the ultimate *in vivo* behaviour of a compound. However, rules based on more information-rich data, such as results from *in silico* models or high throughput assay measurements of physicochemical or biological compound properties, can provide more powerful discrimination. Methods such as rule induction can be applied to

develop and validate multi-parameter optimisation strategies based on these data to more effectively guide the design and selection of high quality compounds in drug discovery. **DDW**

Dr Matthew Segall is CEO and Company Director of Optibrium. Matt has a Master of Science in computation from the University of Oxford and a PhD in theoretical physics from the University of Cambridge. As Associate Director at Camitro (UK), ArQule Inc and then Inpharmatica, he led a team developing predictive ADME models and state-of-the-art intuitive decision-support and visualisation tools for drug discovery. In January 2006, he became responsible for management of Inpharmatica’s ADME business, including experimental ADME services and the StarDrop software platform. Following acquisition of Inpharmatica, Matt became Senior Director responsible for BioFocus DPI’s ADMET division and in 2009 led a management buyout of the StarDrop business to found Optibrium.