Molecular properties that influence oral drug-like behavior
Michael S Lajiness*, Michal Vieth* & Jon Erickson

Address
Structural and Computational Sciences Chemistry Research and Technologies
Lilly Research Laboratories
Eli Lilly & Co
Lilly Corporate Center
Indianapolis
IN 46285
USA
Email: lajinessms@lilly.com; m.vieth@lilly.com

*Correspondence can be addressed to either author

Current Opinion in Drug Discovery & Development 2004 7(4):470-477
© The Thomson Corporation ISSN 1367-6733

Pharmaceutical companies are currently racing to discover the next therapeutic blockbuster. The consensus in the industry is to focus on compounds that are by some measure drug-like, but in order to do this effectively a number of questions must be answered. For example, how should drug-like be defined and how might this definition be used to enhance drug discovery? Has the field moved beyond Lipinski’s seminal ‘rule-of-five’ observations? This review offers a working definition of oral drug-likeness, describes various approaches used in its characterization and discusses its appropriate use. We will focus primarily on the use of computed molecular properties that attempt to predict the oral drug-like behavior of compounds and propose guidelines for the use of observations and trends established in existing datasets to support drug discovery efforts. In particular, this review will demonstrate how trends in simple properties of the data can be used prospectively to compare and prioritize groups of compounds, chemical libraries and different chemical series with greater reliability than for predicting drug-likeness of single compounds. It is the authors’ belief, however, that properties or descriptors that will completely separate drug-like space from non-drug-like space are unlikely to be found; the focus should instead lie on overall distributions of drug-like and non-drug-like compounds in the property space that tend to overlap significantly.

Keywords Bioavailability, bioavailability trends, groups of compounds, molecular properties, oral drug-likeness, oral drugs, predictions, statistical analysis

Introduction
Lipinski’s rule-of-five (ROF) was derived from an analysis of 2245 molecules from the Derwent World Drug Index (DWDI) that were believed to have entered phase II clinical testing [1••]. The rules encompassed by the ROF have been widely implemented throughout the industry and have generally served as a default measure of drug-likeness, even though they were not originally intended for this purpose. Drug-likeness refers to the similarity of a compound to oral drugs and will be referred to as ‘oral drug-likeness’ throughout this review.

Properties that have been associated with oral drug-likeness [2,3,4-6] include:

(i) Oral bioavailability
(ii) Appropriate toxicity to pass phase I clinical trials
(iii) Minimal potency for interacting with a therapeutic target
(iv) Aqueous solubility
(v) Permeability
(vi) Synthetic accessibility
(vii) Pharmacokinetic viability
(viii) Blood-brain barrier permeability (for central nervous system (CNS) drugs)

To proactively apply the above characteristics to drug discovery efforts, computational models would need to be developed for these individual properties. Alternatively, various molecular properties and/or calculated descriptors could be used in their place to identify those features that can discriminate between known oral drugs and presumed non-drugs.

Many researchers have attempted to define operationally ‘drug-like’ by focusing on: (i) simple counting methods [1••,7,8•]; (ii) defining functional group filters [5,6]; (iii) knowledge-based methods [3,9]; (iv) chemistry space methods [10,11]; (v) building blocks [7,12-14,15•,16]; and (vi) other methods [14,17-19].

This review concentrates on the use of relevant molecular properties that describe oral drug-like behavior of compounds, and presents guidelines for the use of observations and trends in existing datasets to support drug discovery efforts.

Defining drug-likeness
Several definitions of oral drug-likeness have been proposed, suggested or inferred in the literature [1••,14,20]. Lipinski defines it as ‘compounds that have sufficiently acceptable ADME properties and sufficiently acceptable toxicity properties to survive through the completion of phase I clinical trials’ [1••,2]. Podolgar et al suggest that drug-like compounds are ‘small molecules, similar to drugs with respect to calculated descriptors that are thought to capture synthetic feasibility, protein binding functionality, and favorable ADME and Tox properties’ [20]. A pragmatic definition by Wang and Ramnarayan states that ‘a compound is drug-like if all its atoms or groups are situated in (molecular) environments similar to those of existing drugs’ [14].

Kubinyi suggests that it is appropriate to select or synthesize compounds that display properties similar to oral drugs but warns that too much emphasis on drug-likeness leads to an overemphasis of ADME data [21]. A common misunderstanding, Kubinyi asserts, is the assumption that ‘molecules that possess drug-like properties and fulfill the Lipinski ROF conditions, are automatically drug-like’. He states that ‘many chemicals conform to these rules, but they are by no means drug-like’.

Muegge notes that drugs typically have between one and three pharmacophore points, and non-drugs are usually under-functionalized (also known as the Ariens and Farmers...
hypothesis [22]) [3•]. These pharmacophore points include groups such as amines, amides, alcohols, ketones, sulfones, sulfonamides, carboxylic acids, carbamates, guanidines, amides, ureas and esters. Hann et al demonstrated that an optimal complexity exists for typical drugs and leads. Thus, it appears reasonable that the potential for useful biological interactions should also form part of an oral drug-like definition [23•].

A poor therapeutic ratio (efficacy/toxicity) is generally unacceptable in a drug, however, it may be a workable finding in an early lead prior to optimization. Therefore, it is our belief that it is more difficult to incorporate toxicological properties into a definition of drug-likeness than to incorporate pharmacokinetic properties since they are more often associated with activity against a specific target. For example, some approved oncolytics are associated with side effects that would not be acceptable for other indications.

Furthermore, while synthetic feasibility or ease of synthesis might be a valid reason to de-prioritize a compound for testing or purchase, this factor need not necessarily be utilized in a definition of drug-likeness. Many approved drugs are synthetically feasible compounds, since toxic effects were selected for clinical development, but some originated from synthetically challenging natural products. Also, the association of a numeric value with synthetic feasibility is problematic.

In light of these concerns, the authors propose a less restrictive definition of oral drug-likeness that includes reference to common properties exhibited by oral drugs; a compound is considered to be oral drug-like when:

(i) The compound possesses a structure having sufficient functionality to potentially favorably interact with a receptor or enzyme.
(ii) The relevant calculated or experimental molecular property profile of the compound is similar to that of known oral drugs.
(iii) The compound is likely to have acceptable ADME properties.

An appropriate metric determining ‘similar’, ‘relevant’ and ‘acceptable’ will vary depending on the intended indication and project stage. This review focuses on the appropriate use of molecular properties relevant to drug discovery efforts, since molecular complexity [12,13,23•] and ADME properties [20] have been recently summarized.

Lead-like versus drug-like
Drawing on the research of Sneader [24], Hann et al demonstrated that drug candidate leads tend to be smaller than the ultimately developed drug [23•]. In an analysis of 269 lead and drug pairs, it was found that drugs having a mean molecular weight (MW) of 314 Da are, on average, 41.6 Da heavier than the corresponding leads (Figure 1). Most leads that lie in the 200 to 300 Da range tend to ‘grow’ by approximately 40 Da to become drugs (Figure 2). Leads that lie in the 400 to 500 Da range tend to grow by approximately 16 Da, suggesting that leads are on average smaller than their corresponding drugs and the lead growth is smaller when starting with a heavier lead molecule.

![Figure 1. The increase in MW from leads to drugs as a function of MW of the final drug for 269 drugs.](image)

The figure shows the increase in MW when going from leads to drugs as a function of MW of the final drug for the 269 Sneader drug set [23•]. The grand MW mean change (horizontal line), and 95% density ellipse are highlighted. (Adapted with permission from the American Chemical Society and Hann MM, Leach AR, Harper G: Molecular complexity and its impact on the probability of finding leads for drug discovery. J Chem Inf Comput Sci (2001) 41(3):856-864. © American Chemical Society.)

In addition, Hann et al demonstrated that leads have fewer heavy atoms, lower cLogP values, fewer aromatic rings, fewer hydrogen bond acceptors and lower Andrew's binding energy, and thus are generally smaller and less complex than their corresponding drugs [9,23•].

In an analysis of the properties of good lead molecules, Oprea [8•] suggested that compounds having MWs not exceeding 450 Da, cLogP values between 4.5 and 3.5, a maximum of four rings, a maximum of ten non-terminal single bonds, no more than five hydrogen-bond donors, and a maximum of eight hydrogen-bond acceptors tend to provide better starting points.

Defining strict cut-off property values as filters for selecting molecules as leads can also cause problems. A number of researchers have highlighted the difficulties of assembling libraries of sufficient size to support high-throughput screening (HTS) without including molecules that venture into a non-drug-like property space [8•,21]. A balance is required in the selection of compounds for screening or synthesis. One solution is to consider the overall drug-like distribution of a library rather than to set hard cut-off values of drug- or lead-like compounds.

Observations derived from compound databases
Several researchers have examined the molecular properties of orally administered drugs and established interesting trends. Wenlock et al [25••] and Blake [26] analyzed the properties of orally administered drugs in various stages of development. The research of Wenlock et al is especially
significant as their study was one of the first to use simple statistical tests for detecting property differences between various groups of clinical compounds. It was observed that the mean properties of compounds in development converge at each stage toward the mean properties of marketed oral drugs (Figure 3) leading to the conclusion that compounds having higher lipophilicity tend to fail in development. These findings suggest that clinical candidates having molecular properties close to the mean of marketed oral drugs tend to be more successful. In addition it was found that trends between drug-likeliness and mean molecular properties are not coincidental but are indeed controlled by the requirements demanded by physiological conditions. The observations of Vieth and co-workers [15••] demonstrating that the property means of US Food and Drug Agency (FDA)-approved drugs have not significantly changed over the past 20 years (eg, see Figure 4) reinforce the conclusions of Wenlock et al [25••]. In contrast to these studies, analyses of clinical candidates from Merck & Co and Pfizer Inc reveal an upward trend in size and lipophilicity [2], highlighting the need for continual monitoring of the relationship between selecting lead molecules biased toward oral drug-like profiles with lower attrition rates and reducing development costs.

Another important contribution to the field was made by Veber et al [27••] with the analysis of trends in the properties of an internal set of GlaxoSmithKline plc (GSK) compounds having measured rat bioavailability and permeation data. Based on a set of 1117 compounds, the observation was made that low bioavailability (< 20%) was, on average, present in compounds with high polarity (expressed as a hydrogen bond donor and acceptor count of > 12 or polar surface area (PSA) of > 140 Å²) and high flexibility (> ten rotatable bonds).

**Deriving rules or generalizations from observations and trends**

Lipinski was arguably the first to indicate trends in simple properties and the absorption of known drugs. The ROF has
Figure 3. Trend in the average MW for drugs in various stages of development.

PI phase I; DI discontinued from development after phase I; PII phase II; DII discontinued from development after phase II; PIII phase III; DIII discontinued from development after phase III; Prereg preregistration; Marketed marketed drugs [25+].

Figure 4. The median MW of FDA-approved oral drugs as a function of approval year.

For each approval year the median MW is shown as a bold horizontal line. Each circle represents a single drug [15+].

had an immense influence on how medicinal chemists expand their corporate compound collections and has inspired research on the molecular properties associated with oral drug-like behavior. Even though the ROF was based on the 90th quantiles of property distributions for a selected group of compounds, it is generally used to classify
single compounds into pass or fail categories. Despite the fact that the Lipinski training set included no 'negative' data, the rules work surprisingly well. There are currently only two marketed drugs that fail all four rules (one oral and one injectable), and 80% of marketed oral drugs pass all four rules [15]. This is largely due to the fact that components of the ROF, namely cLogP and oxygen plus nitrogen (O+N) counts, are anticorrelated, making it difficult to violate the rule, i.e. cLogP > 5 and O+N count > 5. Despite not being derived from more sophisticated statistical methods, such as recursive partitioning or regression, the ROF remains a generally accepted guideline for evaluating the drug-likeness of compounds within the pharmaceutical industry.

Rules derived by Veber et al. [27] have made drug researchers wary of compounds having a high PSA, building on the earlier validation studies of Palm et al [28] and Clark [29]. The resulting 'Veber rules' were based on simple observations, similarly to the ROF, although the validity of the rules should be evaluated for other datasets. For example, in a companion dataset of 277 drugs studied by Sneader (supplied by Veber), the majority of compounds have a molecular similarity that lies within the rules. In addition, the distribution of PSA is different from that of the internal GSK dataset used by Veber et al to formulate the cut-off properties. Consequently, these rules are unable to discriminate between compounds with low measured bioavailability in the Sneader dataset. In the original Veber dataset of 1117 GSK compounds, 44% have high oral rat bioavailability (> 20%), while in the companion Sneader 277 compounds dataset, 84% have high bioavailability. The prediction of bioavailability categories is presented in Table 1. The Veber rules predict correctly that approximately 85% of compounds from the Sneader dataset would have high bioavailability, however, the individual predictions for the compounds of low bioavailability are not as accurate. This observation is indicative of good applicability of the rules to classifying the overall characteristic of the dataset (as of high bioavailability), but highlights the difficulty in classifying the low-bioavailability compounds (the rules correctly predict only 30% of the compounds with low bioavailability). The generalization of group trends works well as an indication of the property of another group in this instance, but caution should be taken when using the rules for single compounds.

**Trends or artifacts?**

Trends in oral drugs are formed on the basis of observations made from a relatively small (in terms of chemical space) finite dataset, while the space in which non-oral drug-like compounds reside is almost limitless. Are molecular property trends observed from the analysis of these small datasets generally applicable or are they an artifact of biased samples? Given the presumed structural diversity of the universe of non-oral drug-like compounds, it appears that compounds sampled from this space would necessarily not cover all property space, including that covered by oral drug-like compounds. Thus, the properties/descriptors that completely separate oral drug-like space from non-drug-like space cannot be uniquely located, making it difficult to predict drug-likeness for any particular compound. It should be possible, however, to demonstrate a difference in the distributions (with respect to location or dispersion) of drug- and non-drug properties, even though the distributions could overlap significantly.

**Table 1. Prediction of the bioavailability (%F) category of the Sneader [24] dataset using Veber rules [27].**

<table>
<thead>
<tr>
<th>Prediction measure</th>
<th>Observed ratio</th>
<th>Correct prediction fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(%F = high</td>
<td>Rule = good)</td>
<td>207/238</td>
</tr>
<tr>
<td>P(%F = low</td>
<td>Rule = bad)</td>
<td>12/39</td>
</tr>
<tr>
<td>P(Rule = good</td>
<td>%F = high)</td>
<td>207/234</td>
</tr>
<tr>
<td>P(Rule = bad</td>
<td>%F = low)</td>
<td>12/43</td>
</tr>
</tbody>
</table>

P(%F = high | Rule = good) is the probability that the observed bioavailability is high given that the Veber rule predicts good bioavailability; P(%F = low | Rule = bad) is the probability that the observed bioavailability is low given that the Veber rule predicts bad bioavailability; P(Rule = good | %F = high) is the probability that the Veber rule predicts good bioavailability given that observed bioavailability is high; and P(Rule = bad | %F = low) is the probability that the Veber rule predicts bad bioavailability given that observed bioavailability is low. The Matthews coefficient, which accounts for the over- and under-predictions of bioavailability, is low at 0.17, but is higher than that expected for the random situation. The calculated probability is based on the observed cell ratios obtained from the classification of the 277 Sneader drug set on bioavailability (high or low) versus Veber Rule prediction (good or bad).

**Validating drug-likeness models**

How can the models used to discriminate between oral drug and non-drug-like compounds be validated? This problem is complex and there is no obvious easy solution. The number of entities considered to be oral drug-like compounds and that have progressed through clinical trials is limited. Are these relatively few compounds representative of the universe of oral drug-like compounds? Lipinski estimates that there are currently only 10,000 drug-like compounds in existence [2]. So, the salient question is whether the compounds used to model oral drug-like behavior are adequate.

Another consideration is the statistical analyses that are applied to derive patterns from historical data. Simple tools, such as means, quartiles and cross tabulation tables, are useful but might miss subtle differences that could lead to erroneous conclusions. Issues concerning the use of experimental design in order to detect differences between groups are often complex requiring expert statistical methods. The use of appropriate statistical tools is beyond the scope of this review, but the authors suggest that statistical guidance should be sought when investigating patterns in historical data.

**Concerns**

**Lack of supporting information**

It is unfortunate that a majority of original observations made by researchers are not supported by publication of the datasets used for the analysis. In only a few cases were data supplied as supporting information (Veber [27] and Vieth [15]), providing other scientists with the opportunity to independently validate findings. It is important to realize that the value of each contribution is enhanced if others can reproduce the results by examining the same dataset.
Choice of datasets to represent oral drug-like or non-drug-like behavior

Much has been published on the molecular properties that influence oral drug-like behavior. While some researchers have utilized opinions based on chemistry in describing compounds that are oral drug-like or non-drug-like [10,30], many have derived their conclusions from analysis of databases that contain oral drug-like compounds (eg, MDL Drug Data Report (MDDR), MDL Comprehensive Medicinal Chemistry (CMC), DWDI [31-33]) and presumably non-drug-like compounds (eg, MDL Available Chemicals Directory (ACD) [34]). In addition, some authors have analyzed internal company proprietary collections and data to draw conclusions [2,27]. Care should be taken in the extrapolation of results obtained from these databases.

As highlighted by Kubinyi, databases such as ACD are not necessarily representative of non-drugs since there are many oral drug-like molecules in that collection [21]. In addition, collections that contain many drug-like molecules (eg, MDDR) contain a large number of biologically inactive compounds. Choosing appropriate subsets of these databases for comparison of drug-like and non-drug-like compounds can potentially bias conclusions. Another issue is that these drug databases contain compounds that have been studied and/or approved over a period of many years. Whether it is suitable to include compounds that were approved 20 years ago could be debatable if the goal is to derive rules that will enhance future new drug application (NDA) submissions. This concern, however, does not appear to be as important for the means of some molecular properties which have not changed over 20 years for FDA-approved drugs [15*]. In addition, proprietary collections might be biased, and generalizations made on the basis of these collections could be problematic. This bias results from historical lead optimization efforts focused around particular chemical classes, such as steroids or benzodiazepines.

Individual versus group comparisons

Often, rules developed to predict oral drug-like behavior are used to assess individual compounds even though this might be an inappropriate use of the trends; this is demonstrated as a comparison of oral and injectable drugs (non-oral-like). Since the distributions of many of the molecular properties for oral and injectable drugs overlap significantly (Figure 5), it is extremely difficult to categorize a compound with any statistical confidence. To further illustrate this point, a simple experiment was carried out using data for oral and injectable compounds [15*]. It was found that the chance of detecting statistically significant differences between oral and injectable compounds increases with sample size (Figure 6). These results demonstrate that it is more likely that differences will be

Figure 5. Distributions of molecular properties for oral versus injectable drugs.

(A) The probability density distribution (P(MW)) of oral (dark) and injectable (light) compounds that fall in selected MW bins. Considerable overlap is observed for the distributions of MW for oral and injectable compounds. (B) The probability density distribution (P(score)) of the oral (dark) and injectable (light) compounds across the four Lipinski score categories. Considerable overlap is observed for the Lipinski score for both types of drug. This is a classic example of group means being statistically significantly well separated, although the distributions do not lead to any useful discriminator.
Figure 6. Success in detecting significant differences in the route of administration of drugs with MW as a function of testing set size.

Success (%) is determined on the basis of the number of statistically significant differences detected between oral and injectable compounds with MW as a function of test-set size. Compounds were chosen randomly 100 times from the set of oral drugs in increasing sample sizes and compared to the total population of injectable drugs, and the p-value for the resultant t-test was obtained. Experiment-wise error (across both oral and injectable drugs) was used as the estimate of variance. In the case of a sample of one, a single compound was randomly chosen from the oral drug population and compared to the entire population of injectable drugs. A t-test was then performed using the variance of the injectable drugs as the common measure of variation. The percentage of statistically significant (< 0.05) differences was plotted versus the number of compounds sampled from the oral drug population.

detected when considering groups of compounds rather than individuals. This notion could have a significant impact within the pharmaceutical industry since collections of compounds (eg, libraries or different chemotype structure-activity relationships (SARs)) are often compared and contrasted in an attempt to prioritize compounds for testing or synthesis. The group effect also holds true for the rank ordering of large compound libraries.

Conclusions and recommendations

Beginning with Lipinski’s ROE, much has been written on the properties associated with ‘drug-likeness’. The motivation for the bulk of this research lies in the desire to be able to use these properties to focus on molecules having the highest probability of lying in the elusive ‘drug-like’ space. In this review, we have attempted to capture these ideas to formulate a practical definition of ‘oral drug-likeness’, namely those compounds that are able to interact (ie, not react) [35] with a receptor, display property profiles of known drugs and have good ADME properties. There is, however, a difference between the properties (eg, MW, clogP, etc) of the final drugs and their leads, which will vary depending on the initial lead. Furthermore, it has been demonstrated that when comparing property distributions of oral drugs and non-oral drugs, the differences can often be subtle. Taking into account these considerations, it is not surprising that it is difficult to make accurate predictions for individual compounds. In fact the use of statistical tests on the differences in property distributions across databases with ranges of oral drug-like character demonstrates that no significant differences are apparent between individual members [36]. Thus, it is recommended that, whenever possible, group predictions should be made in order to maximize the statistical power in selection of the most oral drug-like compounds. Finally, in the rapidly changing drug discovery field where investigators are looking at a multitude of datasets and variables, it is critical to provide supplemental datasets so that others can critically analyze and potentially expand on their full importance and relevance.

Acknowledgments

The authors would like to thank Mike Hann (GlaxoSmithKline) and Daniel Veber (independent drug discovery consultant) for kindly sharing their data. Phil Hipkins and Richard Lewis are thanked for their critical review of the manuscript, and John Rose and Richard E Higgs for fruitful discussions.

References

• of outstanding interest
• of special interest
• This paper is the basis for Lipinski’s ROE, which is widely used in the pharmaceutical industry, and contains a good discussion of solubility and permeability of known drugs. This classic paper should be considered required reading for anyone involved in drug discovery.
• This review surveys computational techniques that are used to evaluate the drug-likeness of compound selections and offers an outlook for further development of the field.


