ACS Medicinal Chemistry Letters

Viewpoint

pubs.acs.org/acsmedchemlett

Improving the Plausibility of Success with Inefficient Metrics

Michael D. Shultz*

Novartis Institutes for Biomedical Research, Inc., 250 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States

ABSTRACT: To increase the probability of success in drug discovery, the concept of drug-like properties was introduced. Efficiency metrics that normalize potency against these properties could help reach drug-like space more efficiently. Potential reasons for the inefficient use of metrics and suboptimal decision making are discussed.

KEYWORDS: *Ligand efficiency (LE), lipophilic efficiency (LipE), drug-likeness, efficiency metrics, decision making, fragment-based drug design*

Let us begin with a self-test. Which of the following is the most probable description of an oral drug?

A. IC_{50} > 10,000 nM.

B. IC_{50} < 1 nM and clogP > 4.

C. IC_{50} = 0.1−1 nM, MW = 450−550, and clogP = 4−6.

This question was presented to 62 drug discovery scientists, and only 7% were able to answer this question correctly based on published data for 258 oral drugs. Two additional questions[1] were asked in order to determine if the pool of respondents were more rational than nonscientists for general probability questions. This data strongly suggests that despite many years of training and experience, drug discovery scientists are irrational in highly predictable ways.

Lipinski et al. helped usher in a new era of property-focused medicinal chemistry to more efficiently identify drug-like molecules. The natural consequence was the introduction and popularization of efficiency metrics, which allegedly provide a quantitative assessment of how much potency is obtained at a cost of another drug-like property. The first such efficiency metric, ligand efficiency (LE), was proposed as "...a way of normalizing the potency and MW of a compound to provide a useful comparison between compounds with a range of MWs and activities."[2] LE is most frequently defined as potency per heavy atom (HA): LE = 1.4pIC_{50}/#HA. Additional metrics have been proposed as alternate ways to normalize potency for additional characteristics such as lipophilicity, polar surface area, enthalpy of binding and sp^3 carbons. Multiple publications link these calculated or measured attributes with various ADMET predictors or outcomes and argue that using these metrics should increase the probability of success in drug discovery.[3]

PROBABILITY, PLAUSIBILITY, POSSIBILITY, AND PROVABILITY

Behavioral economics is a framework for understanding why human psychology leads us to make reliably irrational decisions based on effects such as anchoring, confirmation bias, loss aversion, etc., and can explain why we overestimate our ability to judge probabilities.[4] Kahneman et al. identified attribute substitution, the unconscious substitution of a more difficult problem with an easier problem, as a negative influence on decision making. Assessment of probabilities is an inherently difficult task that requires an accurate understanding of information such as base rate frequencies. By contrast, plausibility is readily assessed because specific, seemingly relevant and easily retrieved information enables rapid heuristic judgments. For the question above, the vast majority of respondents chose C, an answer that seems more plausible, due to the inclusion of an additional, but useless, MW description. This information possibly triggered facile retrieval of the 'rules of 5' and prevented 82% of the respondents from recognizing that as a subset of B, C cannot possibly be more probable (the correct answer is A). Because the assessment of plausibility has greater cognitive ease than probabilistic determinations, attribute substitution often goes unnoticed.

Ligand efficiency, despite being cited approximately 800 times and widely used to normalize potency for size, does not, in fact, normalize potency for size.[5] LE decreases and appears to plateau as size, or number of heavy atoms, increases. Several plausible hypotheses were proposed to explain this observation in terms of ligand flexibility and/or entropic penalties, reduced surface area available for interaction, target specific restrictions, and size-dependent complexity that reduced the probability of optimal fitting. The dependence of LE on #HA is demonstrated utilizing data from the Novartis TNKS program in Figure 1a. This graphical representation of LE (y) as a function of #HA (x) is a $y = m/x$ plot. Therefore, as #HA (x) approaches zero, LE (y) approaches infinity and becomes independent of the actual potency of the compound. If #HA was subtracted from pIC_{50}, a linear relationship would be observed (Figure 1b). Figure 1c demonstrates how lipophilic efficiency (LipE = pIC_{50} − logP) would look if it was calculated as pIC_{50}/logP. All metrics claim to normalize potency against another property, but with the exception of LipE, all other metrics violate the quotient rule of logarithms.[5] As written, they appear plausible but are mathematical impossibilities. Two well-known medicinal chemistry bloggers discussed this viewpoint[6] recognized the flaws of LE and then explained to their readers why they would *continue* to use LE because of the importance of MW. What can we learn from this irrational perseverance?

LIPINKSI'S ANCHOR

The authors of the 'rules of 5' were keenly aware of their target audience (medicinal chemists) and "deliberately excluded equations and regression coefficients...at the expense of a loss

Published: November 21, 2013

   dx.doi.org/10.1021/ml4004638 | *ACS Med. Chem. Lett.* 2014, 5, 2−5

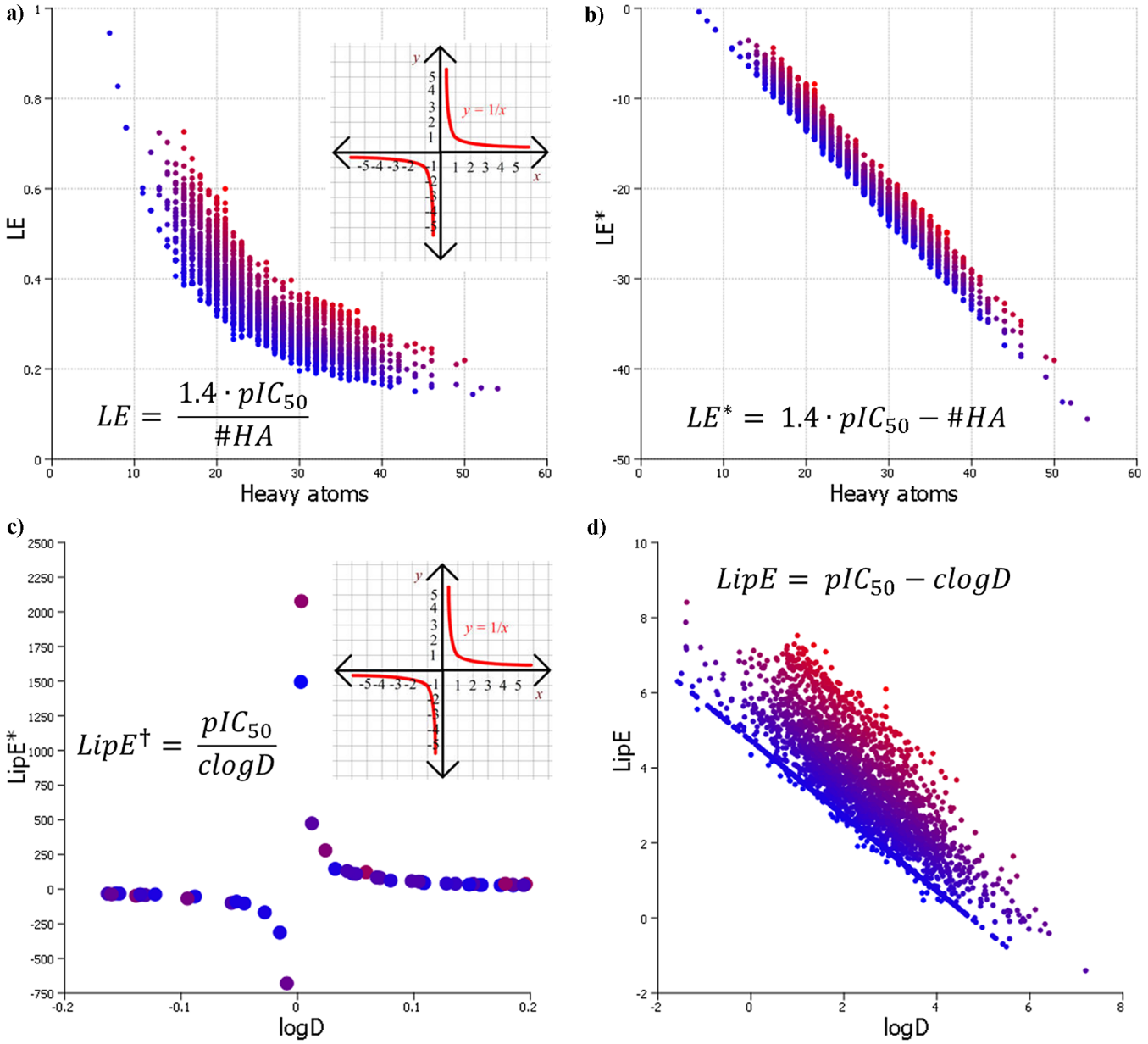

Figure 1. Graphical representation of efficiency metrics for TNKS inhibitors as a function of the number of heavy atoms or clogD. (A) Plot of LE, calculated by dividing pIC_{50} by #HA. If m = 1.4pIC50 and x = #HA, then y = LE = m/x. (B) Plot of LE*, where #HA is subtracted from pIC_{50} ($y = m - x$). Data is colored by potency (blue > 30 μM; red = 1 nM) to highlight lack of size dependency on potency. (C) Plot of LipE†, calculated by dividing pIC_{50} by clogD. If $m = pIC_{50}$ and x = clogD, then y = LipE† = m/x. In contrast to HAC, logD can be negative, and extremely high positive or negative values of LipE† would result as logD values approach zero. (D) Plot of LipE, where clogD is subtracted from pIC_{50} ($y = m - x$).

of detail." One of the greatest misinterpretations of this paper was that these alerts were for drug-likeness. The authors examined the World Drug Index (WDI) and applied several filters to identify 2245 drugs that had at least entered phase II clinical development. Applying a roughly 90% cutoff for property distribution, the authors identified four parameters (MW, logP, hydrogen bond donors, and hydrogen bond acceptors) that were hypothesized to influence solubility and permeability based on their difference from the remainder of the WDI. When judging probability, people rely on representativeness heuristics (a description that sounds highly plausible), while base-rate frequency is often ignored.[4] When proposing oral drug-like properties, the Gaussian distribution of properties was believed, de facto, to represent the ability to achieve oral bioavailability. An anchoring effect is when a number is considered before estimating an unknown value and the original number significantly influences future estimates.[4] When a simple, specific, and plausible MW of 500 was given as cutoff for oral drugs, this became the mother of all medicinal chemistry anchors.

In an apparent affirmation of the importance of MW, it was reported that the average MW of oral drugs in each stage of clinical development decreased until converging on the average MW for oral drugs.[7] The authors concluded this attrition is due to lower drug-like properties of the higher MW compounds earlier in the development pipeline. Plausible? Yes, but worth further examination. An alternate explanation is that compounds in earlier stages of development have different targets than historical drugs (e.g., Bcl-2) and suitable oral bioavailability (e.g., ABT-263) but only more recently entered the

dx.doi.org/10.1021/ml4004638 | ACS Med. Chem. Lett. 2014, 5, 2−5

Table 1. Comparison of Hypothetical (Plausible) Descriptions of Compounds at Various Stages of Drug Discovery with the Most Probable Description of an Oral Drug

	#HA	IC_{50} (μM)	clogP	LE	LLE_{AT}	LipE	LELP
fragment[a]	12	2300	1.0	0.3	0.3	1.7	3.3
hit[a]	20	40	1.6	0.3	0.3	2.8	5.3
lead[a]	30	0.25	2.4	0.3	0.3	4.2	8
candidate[a]	36	0.01	3.0	0.3	0.3	5	10
drug[b]	**22**	**0.063**	**2.5**	**0.42**	**0.38**	**4.7**	**6.0**
			1.4[c]		**0.44[c]**	**5.8[d]**	**3.3**

[a]Values proposed by Mortenson and Murray. [b]Average values from NRDD 258 oral drugs. [c]Average clogD or NRDD 258 oral. [d]LipE based on pIC_{50} − clogD.

development pipeline (e.g., 2006). Since the average MW of approved oral drugs has been increasing while the failure rate due to PK/biovailability has been decreasing, the hypothesis linking size and bioavailability should be reconsidered.

While Lipinski et al. did compare oral drugs with the remainder of the WDI, drugs that are not oral drugs, the base rate frequency (e.g., MW distribution of nondrugs) was not determined. The average MW of nondrugs lies in between oral and nonoral drugs with substantial overlap. MW cannot discriminate between drugs, nondrugs, or route of administration. When a covariate analysis is performed to isolate the correlation between a particular property and multiple in vitro and in vivo outcomes, there is little correlation between MW and either solubility or bioavailability.[8] Higher MW compounds are actually predicted to have *higher* oral bioavailability than lower MW compounds. IV drugs, which are a more soluble class than oral drugs, have larger MW and lower logD values providing additional evidence that MW is not an independent factor for solubility or drug-likeness. The inflation of correlations appears to extend beyond just MW,[9] but the weight of Lipinski's anchor = 0.5 kDa.

If molecular property filters and rules for drug-likeness have limited, if any utility, why is there such a demand for greater discipline in their use?[3] The introduction of highly plausible hit-, lead-, and drug-like definitions has led to further restrictions that affect decision making. LLE_{AT} was introduced with hypothetical definitions of fragments, hits, leads, and candidates (Table 1, Figure 2) with the previously suggested LE value of 0.3 precisely replicated by these definitions. When higher probability values of size, lipophilicity, and potency are used, the target values shift dramatically. Definitions are changed to fit the hypothesis instead of vice versa. This is of critical importance to our industry as investments in fragment-based drug discovery (FBDD) have been proposed to improve the probability of finding attractive starting points.

REASONING IN CIRCLES

LE and FBDD have a symbiotic relationship: A key principle for FBDD is that small molecules can have high LE and one of the proposed strengths of LE is that it identifies weak, but highly efficient fragments. Circular reasoning aside, LE cannot mathematically compare molecules of different molecular weights and potencies because LE is artificially inflated as #HA decreases (see above).[5] Continued use of LE could be explained by theory-induced blindness: the extreme difficulty in noticing the flaws of a theory you have accepted and used as a tool.[4] There are very good reasons for fragment-based approaches, LE notwithstanding.

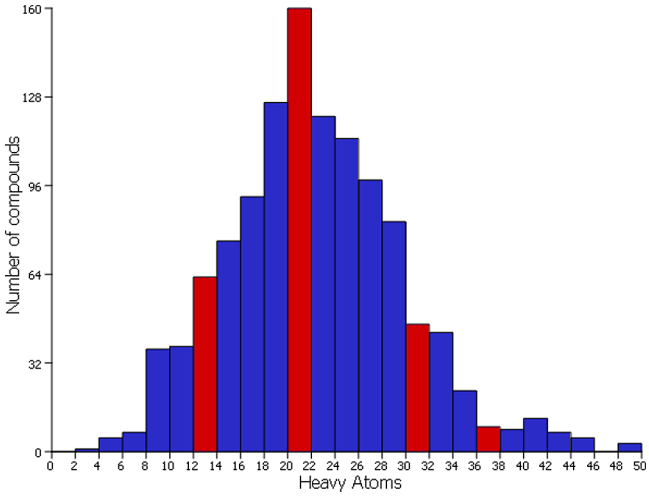

Figure 2. Distribution of 1210 oral drugs as a function of number of heavy atoms (#HA > 50 removed for clarity). Bars marking values of fragments, hits, leads, and candidates, as proposed by Mortenson and Murray, are highlighted in red.

A clear understanding of probabilities in drug discovery is impossible due to the large number of known and unknown variables. Although some metrics may have utility (e.g., LipE and enthalpy),[10] all metrics have limitations.[5] It is important to know when metric-based information is useful, useless, or possibly misleading so it can be prioritized appropriately. The use of metrics should improve the probability, not just the plausibility, of success.

AUTHOR INFORMATION

Corresponding Author

*E-mail: michael.shultz@novartis.com. Phone: +1 617-871-7551.

Notes

Views expressed in this editorial are those of the author and not necessarily the views of the ACS.

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The author acknowledges Young Shin Cho and Barry Touré for helpful discussions and suggestions on this manuscript.

ABBREVIATIONS

LE, ligand efficiency; LipE, lipophilic efficiency; LLE_{AT}, Astex LLE; #HA, number of heavy atoms; WDI, world drug index; HTS, high throughput screen; ADMET, adsorption metabolism; excretion and toxicity; MW, molecular weight; TNKS, tankyrase; FBDD, fragment-based drug discovery

REFERENCES

(1) Question (1). Choose one of the following: D. 85% chance to win $1,000 (15% chance to win nothing) or E. Win $800 for certain. Question (2). Choose one of the following: F. 85% chance to lose $1,000 (15% chance to lose nothing) or G. Lose $800 for certain. Option E was chosen by 78%, and option F was chosen by 87% of the respondents, despite being the lower value and less rational choices.

(2) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* **2004**, *9*, 430−431.

(3) Johnstone, C. Medicinal chemistry matters: a call for discipline in our discipline. *Drug Discovery Today* **2012**, *17*, 538−543.
(4) Kahneman, D. *Thinking, Fast and Slow*; Farrar, Straus and Giroux: New York, 2011; p 499.
(5) Shultz, M. D. Setting expectations in molecular optimizations: Strengths and limitations of commonly used composite parameters. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 5980−5991.
(6) Myriad Metrics − but Which Are Useful? http://practicalfragments.blogspot.com/2013/08/myriad-metrics-but-which-are-useful.html, Too Many Metrics http://pipeline.corante.com/archives/2013/08/22/too_many_metrics.php, More Thoughts on Compound Metrics http://pipeline.corante.com/archives/2013/09/04/more_thoughts_on_compound_metrics.php.
(7) Wenlock, M. C.; Austin, R. P.; Barton, P.; Davis, A. M.; Leeson, P. D. A comparison of physiochemical property profiles of development and marketed oral drugs. *J. Med. Chem.* **2003**, *46*, 1250−1256.
(8) Sutherland, J. J.; Raymond, J. W.; Stevens, J. L.; Baker, T. K.; Watson, D. E. Relating molecular properties and in vitro assay results to in vivo drug disposition and toxicity outcomes. *J. Med. Chem.* **2012**, *55*, 6455−6466.
(9) Kenny, P. W.; Montanari, C. A. Inflation of correlation in the pursuit of drug-likeness. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 1−13.
(10) Shultz, M. D. The thermodynamic basis for the use of lipophilic efficiency (LipE) in enthalpic optimizations. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 5992−6000.