J Comput Aided Mol Des (2014) 28:699–710
DOI 10.1007/s10822-014-9757-8

PERSPECTIVE

Ligand efficiency metrics considered harmful

Peter W. Kenny · Andrei Leitão · Carlos A. Montanari

Received: 20 March 2014 / Accepted: 26 May 2014 / Published online: 5 June 2014

Abstract Ligand efficiency metrics are used in drug discovery to normalize biological activity or affinity with respect to physicochemical properties such as lipophilicity and molecular size. This Perspective provides an overview of ligand efficiency metrics and summarizes thermodynamics of protein–ligand binding. Different classes of ligand efficiency metric are critically examined and the study concludes with suggestions for alternative ways to account for physicochemical properties when prioritizing and optimizing leads.

Keywords Acid/base properties · Ligand efficiency · Lipophilic efficiency · Metric · Property-based design

Introduction

Ligand efficiency metrics (LEMs) are used in pharmaceutical molecular design to normalize biological activity or affinity with respect to physicochemical properties such as lipophilicity and molecular size which are widely assumed to be important determinants of in vivo behavior of drugs [1]. Despite their wide acceptance and use by the drug discovery community, the physicochemical basis of these metrics is weak and their relevance and validity have been challenged recently [2–4]. We first provide an overview in which we articulate assumptions commonly made when defining LEMs. Following a short summary of ligand binding thermodynamics, we critically review different types of LEM before recommending alternatives to LEMs for normalizing activity with respect to physicochemical properties when ranking and optimizing leads. In particular we suggest using the trend that is actually observed in the activity data to normalize it rather than some arbitrarily assumed trend.

Given the focus of this journal, it is appropriate to first place efficiency metrics within the molecular design context. Molecular design can be defined as control of the behavior of compounds and materials by manipulation of molecular properties [5] and thought of as hypothesis-driven or prediction-driven [6]. There are parallels and a degree of overlap between hypothesis-driven molecular design and statistical molecular design [7] which are frameworks for assembling structure–activity relationships (SARs) as efficiently and systematically as possible. LEMs are elements of property-based design [8, 9], the basis of which is the assumption that physicochemical properties of such as lipophilicity and molecular size are important determinants of pharmacokinetic and toxicological behavior of compounds. When we normalize activity data, we try to quantify the contribution of a physicochemical property to activity so that it can be subtracted from the measured value. Quantifying these contributions requires that assumptions (e.g. linear response) be made about the underlying relationship between biological activity and physicochemical properties. The validity of the normalization procedure is limited by the extent to which these assumptions hold. Normalization transforms activity data and can be seen as a type of quantitative structure activity relationship (QSAR). The justification for normalizing activity depends ultimately on the extent to which the chosen physicochemical properties are predictive of relevant in vivo behavior of compounds. Drug discovery

P. W. Kenny (✉) · A. Leitão · C. A. Montanari
Grupo de em Química Medicinal do IQSC/USP – NEQUIMED, Instituto de Química de São Carlos, Universidade de São Paulo, Av. Trabalhador Sancarlense, 400, São Carlos, SP 13566-590, Brazil
e-mail: pwk.pub.2008@gmail.com

scientists need to be especially wary of data being presented in ways that exaggerate trends and inflate correlations [10] because this can lead to unjustifiably restrictive design guidelines [11] being applied when selecting compounds for screening or synthesis.

The rationale for using LEMs is to allow more relevant comparison of compounds than is possible by affinity or potency alone. It is usually assumed that excessive lipophilicity and molecular size lead to unacceptable pharmacokinetics and toxicology and these physicochemical properties can be considered as risk factors in the context of pharmaceutical molecular design. Affinity with respect to the primary target is also assumed to increase with lipophilicity and molecular size so it is important that compounds be designed to exploit these characteristics as effectively as possible when engaging their targets. An LEM can be thought of as an attempt to quantify the extent to which compounds beat (or are beaten by) a trend in the activity data and this is sometimes expressed colloquially as 'bang for buck'. LEMs are primarily used to compare and rank compounds such as screening hits or those synthesized during the course of optimization. Fragment-based drug discovery (FBDD) [12–19] makes extensive use of LEMs to monitor progress from initial screening hits that may bind with affinity as weak as millimolar and combinations of LEMs have also been used to map chemicobiological space [20, 21]. When considering the use of LEMs, it should be remembered that the primary determinants of target engagement in vivo are free concentration and affinity and not free fraction and ligand efficiency [10, 22].

LEMs are functions of measured activity, for example potency in an enzyme kinetic assay, and one or more risk factors such as lipophilicity or molecular weight. When we use an LEM we make implicit assumptions about the underlying relationship between activity and the physicochemical property or properties with which we choose to quantify risk [2]. Our criticism of currently used LEMs is that defining assumptions are made arbitrarily and take no account of the possibility that different activities might respond differently to a risk factor. LEMs are typically defined by either scaling or offsetting activity with respect to the values of the chosen risk factors although there is no reason that they should be limited to these functional forms [2]. Scaling means that measured activity is divided by the value of the physicochemical property [2]. For example, ligand efficiency (LE) is defined by scaling standard free energy of binding (ΔG°) by number of heavy (i.e. non-hydrogen) atoms and has units of molar energy per heavy atom [23]

$$\mathrm{LE} = -\Delta G^\circ / \mathrm{HA} \quad (1)$$

Drugs typically inhibit the action of their targets and it is pIC_{50}, often referred to as potency, rather than ΔG° that is used more commonly in pharmaceutical design to quantify activity. This quantity is obtained by transforming the half-maximal inhibitory concentration (IC_{50}) where M is the concentration unit of mol/litre.

$$\mathrm{pIC}_{50} = -\log_{10}(\mathrm{IC}_{50}/\mathrm{M}) \quad (2)$$

Offsetting means that the value of the chosen physicochemical property is subtracted from measured activity so it is imperative that both be expressed in the same units [2]. For example, the following was suggested [13] as a criterion for compound being likely to be a good lead where the distribution coefficient, D, is a function of pH.

$$\mathrm{pIC}_{50} - \log \mathrm{D} \geq 2 \quad (3)$$

One criticism that can be made of using LEMs as a framework for drug discovery is that it is rarely, if ever, articulated why activity has been scaled to create one LEM but offset to create another.

Whether we scale or offset activity by a physicochemical property, we are assuming that the underlying relationship between activity and risk factor is linear. Scaling assumes that the relevant line passes through the origin (i.e. zero intercept) while offsetting implies that it has unit slope [2]. Assuming specific values of the slope and intercept leads to systematic normalization errors when the assumptions are incorrect as illustrated in Fig. 1 in which the data points all lie on a straight line with non-zero intercept and non-unit slope. Forcing unit slope or zero intercept for this data set leads to the erroneous conclusion that Y values beat the trend at low X but are beaten by it at high X. Making incorrect assumptions about trends in data

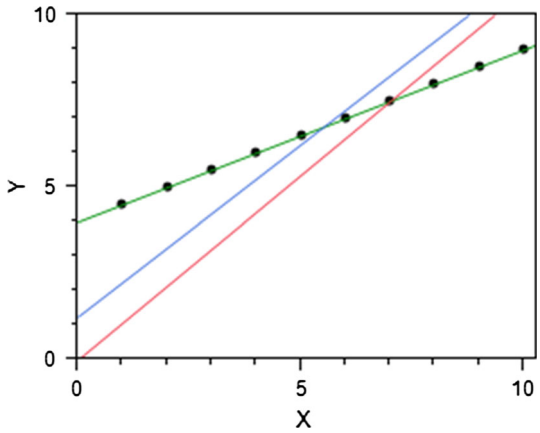

Fig. 1 Artifacts introduced by constraining slope or intercept when fitting linear models. The data points all lie on a line (*green*) with slope of 0.5 and intercept of 4 and the other lines are results of fitting data with constraints of zero intercept (*red*) or unit slope (*blue*). For this data set, the effect of each constraint is to make it appear that Y values with low X values beat the trend while those with high X value are beaten by it

distorts our perceptions whether we are ranking screening hits or depicting lead optimization trajectories.

Drug discovery scientists are presented with what can appear to be a bewildering array of metrics to guide them in their quest for high quality compounds [3] and the casual observer may form the impression that relevance has been sacrificed for simplicity. Although LEMs are touted as quantitative measures of compound quality and might be regarded as simple QSAR models, they are not typically validated [24] in a quantitative manner. One problem is that compound quality, which it is believed that LEMs quantify, is not something that one can measure and even the term itself is not well defined. Evidence presented in support of an LEM is usually indirect and circumstantial. For example, it might be asserted that the LEM has been used to guide medicinal chemistry in a particular project or that the LEM discriminates between compounds believed to be of higher quality (e.g. marketed drugs) and compounds believed to be of lower quality (e.g. high throughput screening hits). When evidence is presented graphically or without objective statistical analysis, it is impossible to know how strong the discrimination actually is and ways in which trends can be exaggerated (e.g. by masking variation with standard error) have already been discussed at length [10]. Graphical representations are useful, even essential, for exploring data prior to analysis but if we are unable to quantify the strengths of trends, we cannot determine objectively whether one metric is more effective than another. Implicit assumptions (e.g. linear model with zero intercept) are always made when defining LEMs but testing how well the assumptions actually hold when applied to real data is rarely, if ever, done in order to validate the metrics. One point that potential users of LEMs need to consider carefully is whether or not they believe that activity measurements lying on a straight line when plotted against risk factor indicate equal efficiency.

Thermodynamics of binding

As observed by Ehrlich over a century ago, drugs exert their actions by binding to their targets (and anti-targets) and a short review of relevant thermodynamics will help to clarify some of the deficiencies of efficiency metrics. The standard free energy of binding (ΔG°) can be written in terms of the dissociation constant (K_d) and the concentration (C°) used to define the standard state.

$$\Delta G^\circ = RT\log_e(K_d/C^\circ) \quad (4)$$

Although C° is usually omitted from Eq. 4 in drug discovery and biochemical literature, it is incorrect to do so because the logarithm function is defined for (dimensionless) numbers but not quantities with units. The chemical potential μ of a solute at a concentration C in a dilute solution can be written as follows, where μ° is the chemical potential at the standard concentration.

$$\mu = \mu^\circ + RT\log_e(C/C^\circ) \quad (5)$$

Equation 4 can be derived from Eq. 5 and ΔG° is simply the difference between standard chemical potential for protein–ligand complex (P.L), ligand (L) and protein (P).

$$\Delta G^\circ = \mu^\circ(P.L) - \mu^\circ(L) - \mu^\circ(P) \quad (6)$$

As noted by Gilson et al. the standard concentration, C°, is arbitrary [25–27] and the standard state does not even have to be physically accessible. Readers might consider a 1 M solution of a protein to convince themselves of the second point. Had 1 nM been adopted as the standard concentration, ΔG° would be positive for a K_d of 10 nM. Since only the entropic term in ΔG° depends on C° (the solution is dilute) it is always possible to set C° so that ΔS° is zero for the formation of a particular protein–ligand complex and this should be kept in mind when asserting that binding of a ligand is driven by enthalpy or by entropy.

The primary assay in a drug discovery project typically measures potency rather than affinity and access to K_d is usually limited. Substituting IC_{50} for K_d in Eq. 4 is neither correct nor necessary and the motivation for doing so may be to make potency measurements appear more 'physical' by endowing them with units of molar energy. Affinity is more useful than potency for developing quantitative models of protein–ligand binding but potency, when measured with substrates and co-factors at physiological concentrations, is likely to be more predictive of target engagement in cellular assays or in vivo. The action of a drug is as much a function of free concentration (C_{free}) as of in vitro potency or affinity and we can think of the ratios C_{free}/IC_{50}, C_{free}/K_d (and their logarithms) as measures of target engagement potential (TEP). Equation 2 shows that pIC_{50} is defined using what we will call a 'reference concentration' and the choice of 1 M to define this is arbitrary just as the choice of C° is arbitrary. Using a reference concentration allows an IC_{50} value to be divided by its units so that the logarithm can be calculated but it must be remembered that the p operator has molarity built into its definition. Even when K_d is available, it can be argued that it would still be better to use $-\log_{10}(K_d/M)$ (pK_d) rather than ΔG° for defining LEMs because the units are often incorrectly discarded when LEMs are presented in drug discovery literature. For example, most of the LE values in the H2014 study [1] are presented as dimensionless quantities. A more subtle example of the problem is provided by the definition of LLE_{AT} [28] as the sum of a (dimensionless) number and a quantity with units of molar energy per heavy atom.

Since the choice of standard (for ΔG°) or reference (for pIC_{50}) concentration is arbitrary [25–27], either the relative values of an LEM for different compounds or the differences between them must, in general, be invariant with respect to the choice of standard or reference concentration in order for the LEM to be valid for comparing compounds. If this were not the case one might infer that compound A was more ligand efficient than B for one value of C° only to come to the opposite conclusion when another value of C° was used. Put another way, changing the units of quantities that characterize a system should not alter our perception of that system. The only situation in which the requirement for invariance could be relaxed would be if a particular value of C° could be proven to be optimal, either on strong theoretical grounds or by relevant data analysis. It is sometimes incorrectly assumed that values of zero for ΔG° or pIC_{50} have a special significance such as indicating an absence of interaction between ligand and protein. A zero value in this context is simply a statement that K_d or IC_{50} is equal to the standard or reference concentration [2] so it is not, in general, correct to constrain lines of fit to pass through the origin when modelling the response of ΔG° or pIC_{50} to molecular size. The slope of a line that has been constrained [29, 30] in this manner is of questionable value as an estimate for the size of the relevant response as is illustrated in Fig. 1. An analogous issue arises when establishing appropriate zero points for solute hydrogen-bonding scales and this has been addressed by analysis of data measured for multiple reference bases and acids [31].

Molecular interactions [32] figure prominently in pharmaceutical molecular design, an activity that often involves searching for new ways to form interactions between ligand and target. In general, the contribution of a particular intermolecular contact (or group of contacts) to affinity (or the changes in enthalpy, entropy, heat capacity or volume associated with binding) cannot be measured experimentally. This means that one needs to be wary of what one of the reviewers of the manuscript described as, "a bolder intention with LEMs, which is to quantify the reference contribution a certain atom (or atom type) should have on an activity". Studies of relationships between affinity and molecular size suggest that affinity tends to plateau and this is usually interpreted as reflecting a reduced likelihood of forming optimal interactions as ligand molecules become larger [29, 30, 33, 34]. However, other factors may come into play in that very high affinity ligands may neither be sought by nature nor by medicinal chemists [29] and quantifying very high affinity or potency is not always straightforward. For example, the lowest IC_{50} that can be measured routinely in an enzyme kinetic assay is limited by the concentration of enzyme [35]. If we assert that an upper limit to affinity has been found then we also need to show that the limit is not one that the assay has imposed. This issue is not confined to enzyme kinetic assays and upper quantitation limits for permeability can also reflect assay configuration rather than diffusion.

Ligand efficiency and scaled efficiency metrics

Scaling activity by molecular size can be thought of as an attempt to quantify something that, for want of a better term, can be described as intensity of interaction between ligand and target.

Ligand Efficiency (Eq. 1) was the first LEM to be introduced [23] and it was suggested subsequently [15] that pIC_{50}/HA could also be used for this purpose. The binding efficiency index (BEI) and surface binding efficiency index (SEI) for which activity is scaled by molecular weight (MW) or polar surface area (PSA) have also been defined [36].

$$\mathrm{BEI} = (\mathrm{pK_i \text{ or } pK_d \text{ or } pIC_{50}})/\mathrm{MW} \quad (7)$$

$$\mathrm{SEI} = (\mathrm{pK_i \text{ or } pK_d \text{ or } pIC_{50}})/\mathrm{PSA} \quad (8)$$

Activity can also be scaled by molecular surface area [15] and new metrics have been created by changing units of concentration [37]. The scope of activity has been broadened to include binding enthalpy [38], metabolic stability [39] and residence time [40].

Uncertainty about the appropriate zero molecular size limit for activity is the main problem when defining LEMs by scaling. The arbitrary nature of the standard concentration used to define ΔG° and the reference concentration used to define pIC_{50} means that it is incorrect to simply assume zero values of these to be the appropriate zero molecular size limits. It needs to be noted that this error has been made in an article [29] that is widely-regarded as providing the theoretical foundation for ligand efficiency [1, 23, 30]. The effect on ligand efficiency of changing the standard concentration is shown in Table 1 and it should be noted that the LEM used in this illustrative example has been defined in a manner to make it unitless. When C° is set to 1 M the three compounds with 10, 20 and 30 heavy atoms would be regarded as equally ligand-efficient. Whilst efficiency increases with number of heavy atoms when C° is set to 0.1 M the opposite behavior is observed when C° is 10 M. Fig. 2 shows how three parallel lines transform when the Y coordinate is divided by the X coordinate. The degree of curvature reflects the ratio of intercept to slope and is most marked for low values of X which has clear implications for FBDD [12–19] where numbers of heavy atoms are relatively low. The results shown in Table 1 and Fig. 2 suggest that some of the size dependency [34] of LE may be simply an artifact of the arbitrary choice of 1 M as the reference or standard concentration.

Table 1 Effect of standard concentration on ligand efficiency

N^a_{HA}	K_d/M^b	$C°/M^c$	$-(1/N_{HA}) \times \log_{10}(K_d/C°)^d$
10	10^{-3}	1	0.30
20	10^{-6}	1	0.30
30	10^{-9}	1	0.30
10	10^{-3}	0.1	0.20
20	10^{-6}	0.1	0.25
30	10^{-9}	0.1	0.27
10	10^{-3}	10	0.40
20	10^{-6}	10	0.35
30	10^{-9}	10	0.33

[a] Number of heavy atoms (dimensionless)

[b] Dissociation constant

[c] Standard concentration

[d] The ligand efficiency metric has been defined, for clarity, to be unitless by using pK_d instead of $\Delta G°$ and N_{HA} instead of HA

A more fundamental question, however, is not what value we should use for the zero molecular size limit but whether a single value is appropriate to all assays. Consider an arbitrary selection of ATP-competitive inhibitors of a generic tyrosine kinase for which pIC_{50} is measured at two different ATP concentrations. This might have been done because the screening assay (run at lower ATP concentration to make it easier to observe inhibition) is less predictive of activity in cells where ATP concentration is higher. Would we expect the same zero molecular size limit to apply for both sets of assay results? Even when affinity is used rather than potency, different responses to molecular size will still, in general, be observed for different targets and structural series. An inhibitor deconstruction study [14] goes some way to addressing this question and the study data, consisting of affinity/potency measurements for minimal binding elements on a number of optimization trajectories is plotted in Fig. 3. This particular data set is organized by target, making it relevant for the drug discovery scientist who is primarily interested in how the ligand efficiency concept might be used to guide decision-making in a specific project. The intercepts resulting from fitting linear models to the data for each target range from −1.71 to + 4.69 which supports the hypothesis that different zero molecular size limits for activity apply to different assays. This analysis calls into question the validity of combining results from different assays when using LEMs to map chemico-biological space [20, 21] or to explore the size dependency of LE [34]. This data set also shows how maximal activity may appear to plateau even when linear responses are observed for individual assays and this is especially relevant in the context of correcting LE values for the effects of molecular size [34]. When asserting that activity tends to plateau with molecular size, it is important to demonstrate that this is also the case for individual assays and not just for the full data set.

Others have noted that to scale activity is to assume that $\Delta G°$ is zero in the limit of zero molecular size although they do not appear aware of the work of Gilson et al. [25–27], Abraham [31] or a recent study [2] in which this issue was discussed specifically in the context of ligand efficiency. The H2014 study states,

> Varying the temperature and standard concentration will change the relative ΔG [1].
>
> Thus for a given ligand efficiency (the gradient) there is a linear relationship between the free energy of binding and the number of heavy atoms, where it is assumed that a ligand with zero atoms has zero free energy of binding at standard conditions [1].

The H2014 study [1] fails to make the connection between the two statements and does not question the validity of assuming zero free energy of binding for a ligand with zero atoms. In contrast, the S2013 study [3] does demonstrate awareness that it is incorrect to force the line through the origin and is prepared to criticize, as we have also done, a study [29] that is widely regarded [1, 23, 30] as providing

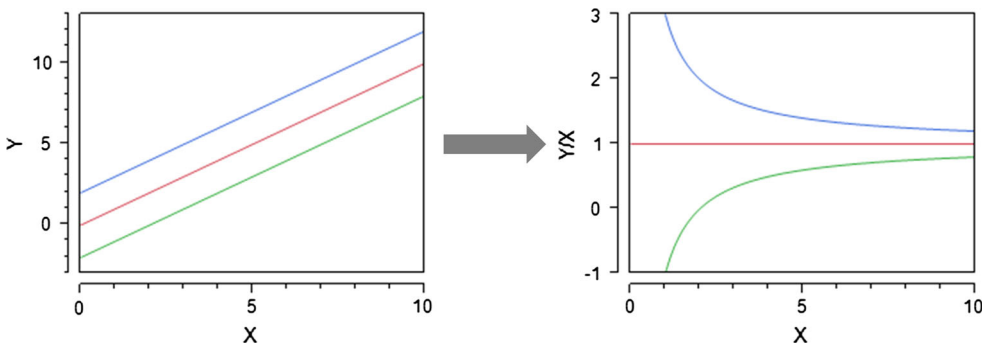

Fig. 2 Effect of scaling transformation on *three parallel lines* of unit slope. The (*red*) line passing through the origin transforms to a line of zero slope while the other lines transform to *curves*. The curvature of a line transformed in this manner reflects the sign of the intercept and the relative magnitudes of the intercept and slope of the line in question

the theoretical basis of ligand efficiency. The S2013 study asserts.

> Their conclusion came from the relationship between potency and HAC for selected compounds with fewer than 15 heavy atoms where the slope of the line was 1.5 kcal/mol per heavy atom when the line is forced through identity. A logical fallacy in the form of a circular argument was committed. It can only be justified that the line should go through zero if there is a relationship between HAC and potency, however it was the slope of the line (through zero) that was used to demonstrate the relationship between HAC and potency [3].

The S2013 study argues a case against LE, a key part of which is an assumption that the change in potency per heavy atom required to maintain a given value of LE must be constant for the metric to be valid [3]. This is actually equivalent to requiring that activity values for compounds lie on a straight line with zero intercept. Suppose that the affinities of a number of compounds are given by:

$$-\Delta G^{\circ} = a + (b \times HA) \quad (9)$$

Equation (9) ensures that a given change in HA will always result the same change in affinity and LE can be written as:

$$LE = (a/HA) + b \quad (10)$$

This shows that when affinity is a linear function of HA, LE is only constant when the intercept is zero and this is equivalent to what has already been presented in Table 1 and Fig. 2. The S2013 [3] and S2014 [4] studies assert that LE is mathematically invalid and this view has been challenged [41]. Our view is that LE is mathematically valid but it is incorrect to invoke thermodynamics in its support or to present values of LE as evidence that one compound is better than another.

Approaches have been proposed for presenting values of LEMs graphically [20, 21] although the lack of means with which to compare the relevance and predictive power of different mappings in an objective manner limits their utility. Typically, values of two LEMs are plotted to display the activity of the compounds of interest in an efficiency plane and the philosophy behind the approach can be summarized as,

> Drug discovery should not need to be 'a numbers game' anymore but an efficiency-based process driven by a suitable choice of variables [21].

It can be argued that the process of drug discovery should be based on target engagement potential rather than efficiency and there are a number of reasons why LEMs are not appropriate alternative variables with which to map chemico-biological space. The main problem is that the ordering of compounds by efficiency is dependent on the concentration used to define the standard state and there is no way of knowing a priori whether the conventional value of 1 M is optimal for the purpose of mapping. Furthermore, scaling affinity generally transforms a linear response to molecular size into a non-linear relationship between LE and molecular size. Typically, efficiency planes are defined using two different LEMs (sharing a common term such as pIC_{50} or $-\Delta G^{\circ}$) and this unnecessarily introduces correlation that affects the visual impact of the data representation.

One scaled LEM that does not suffer from the problem of an unknown zero molecular size limit for activity is group efficiency (GE) [42] which is defined using differences between ΔG° measurements. This requires that one compound be selected as a reference and has the disadvantage that uncertainty in differences between ΔG° values is greater than that in the ΔG° values themselves which increases sensitivity to noise in the assay. Uncertainty is an important consideration when comparing values of LEMs and the scaling transformation has another, more subtle effect which is completely unrelated to the issue of an unknown zero molecular size limit. When we scale activity, we also scale the uncertainty in the activity. Consider 0.25 kcal/mol per heavy atom and 0.30 kcal/mol per heavy atom as two values of LE and let us also assume a constant uncertainty in the ΔG° measurement. The uncertainty in LE is threefold higher for a fragment with 10 heavy atoms than for an optimized clinical candidate with 30 heavy atoms which means that we should become less confident that a particular difference in LE values is meaningful as molecules get smaller. The implications for comparing LE values across a wide molecular size range, as would be the case for a fragment-based project, should be obvious.

Lipophilic efficiency and offset efficiency metrics

We will start this section by noting that, in the current study, the base 10 logarithm of any predicted 1-octanol/water partition coefficient will be referred to as ClogP and the alternative capitalization cLogP will not be used. Offsetting activity by lipophilicity can be seen as an attempt to quantify ease of transferring the ligand from a non-polar environment to its binding site in the protein and inequality (3) appears to represent the first published example [13] of an LEM being defined in this manner. It was later suggested that (pIC_{50} − logP) could also be used for this purpose and the use of predicted values of both logP and logD in this context was also discussed [15]. Subsequently, ligand-lipophilicity efficiency (LLE) [43] and lipophilic efficiency (LipE) [44] were named as LEMs. The former has since been renamed [45] as lipophilic ligand efficiency while the latter has been redefined [46] to bring it into line with LLE.

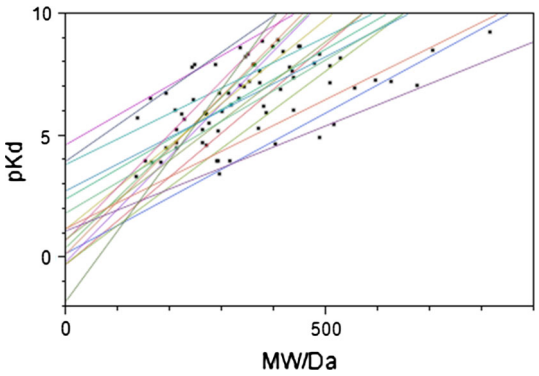

Fig. 3 Plot of pK_d against molecular weight showing variation in intercept with target. Data are from the inhibitor deconstruction study described in ref 14 and pK_d was fit to molecular weight, using a linear model, (pK_d = intercept + slope × MW/Da) for each of 18 targets. No attempt has been made to distinguish data points corresponding to each target because the purpose of this plot is to illustrate the variation of intercept and slope over the 18 targets. The intercept terms (standard error in parenthesis) range from −1.71 (0.25) to +4.69 (0.87) for the individual targets and the intercept for the full data set is +4.27 (0.42). The r^2 values range from 0.845 to 0.999 although the corresponding value for the full data set is only 0.325. Data analysis was performed using JMP 10.0.2 (www.jmp.com, accessed 28 Feb 2014)

$$\text{LLE} = (\text{pIC}_{50} \text{ or } \text{pK}_i) - (\text{logD or ClogP}) \quad (11)$$

$$\text{LipE} = (\text{pIC}_{50} \text{ or } \text{pEC}_{50}) - \text{ClogP} \quad (12)$$

The degree of redundancy between the two definitions is unfortunate although the definition of LLE is particularly unsatisfactory because it does not specify which measure of lipophilicity should be used. For example, there is no way of knowing exactly what "LLE target of ~5–7 or greater" [43] means for a carboxylic acid without seeking further information. The use of predicted logP values raises an analogous, albeit less obvious, issue because different prediction methods (and even different versions of a particular method) will, in general, yield different results for the same molecular structure [47]. Furthermore, it is also necessary to specify pH when using logD. We can legitimately question the value of naming a metric if it becomes necessary to state exactly what is meant by the name each time the metric is used.

Although differences in values of LEMs derived by offsetting activity are at least invariant with respect to the standard or reference concentration, the assumption of unit slope in the defining linear model can still be questioned. For example, we might ask whether (pIC_{50} − 0.5 × ClogP) might not be a better LEM than (pIC_{50} − ClogP) and it is useful to define a generalized lipophilic efficiency, $LipE_{gen}$, to describe a continuum of metrics specified by the coefficient, λ, of lipophilicity.

$$\text{LipE}_{gen} = \text{pIC}_{50} - (\lambda \times \text{ClogP}) \quad (13)$$

This issue is entirely analogous to that of establishing the appropriate zero molecular size limit for a scaled LEM and we can extend the analogy to ask whether or not a single value of λ would be appropriate to every activity that might be measured. The effects of offsetting on straight lines of different slopes are shown in Fig. 4 which shows how this data transformation would lead to a view that lipophilic efficiency is lipophilicity-dependent for any line except that with unit slope.

Although studies have presented results that support the use of lipophilic efficiency in drug discovery, none of these appear to have considered the possibility that the coefficient, λ of lipophilicity in $LipE_{gen}$ should be anything other than unity. The H2014 study asserts that "LLE correlates positively with the drug efficiency index (DEI)" [1]. LLE and DEI (sum of pIC_{50} and base 10 logarithm of dose-scaled free plasma concentration which is actually a measure of TEP) [48] share pIC_{50} as a common term and so it is incorrect to invoke a correlation between the two metrics in support of LLE. A study [49] of clinical VEGFR TK inhibitors presents results that suggest that both potency against the enzyme and lipophilicity influence kinase selectivity and clinical efficacy. However, it is much less clear that LipE (λ = 1) is optimal for quantifying the contributions of these two factors and the strength of the correlation with kinase selectivity owes much to the exclusion of sunitinib. Correlations between LipE and enthalpy of binding have been presented [50] for trypsin, renin and HIV protease although it is not clear whether lipophilic efficiency or binding enthalpy is in greater need of validation as a measure of compound quality. The optimal value of λ is but one of a number of questions facing the drug discovery scientist who wishes to normalize activity with respect to lipophilicity. For example, either logP or logD can be used to define the metric and decisions also may need to be made about prediction method (for ClogP) or pH (for logD). It can be instructive to ask whether evidence presented in support of a particular LEM could have been used to show objectively that the LEM in question had performed more effectively than plausible alternatives.

Although we cannot simply assume that λ is unity we might invoke a thermodynamic interpretation of (pIC_{50} − logP) to justify the functional form of the metric which might be regarded as measure of the ease of transferring the ligand from a non-polar environment to its binding site in the protein. There are two flaws in this line of reasoning. The first problem is that wet 1-octanol may not be the most appropriate reference solvent since it can form hydrogen bonds with solutes and the 1-octanol/water logP partitioning system appears to not sense hydrogen bond donors to the same extent as it senses acceptors [2, 10, 51]. Whilst

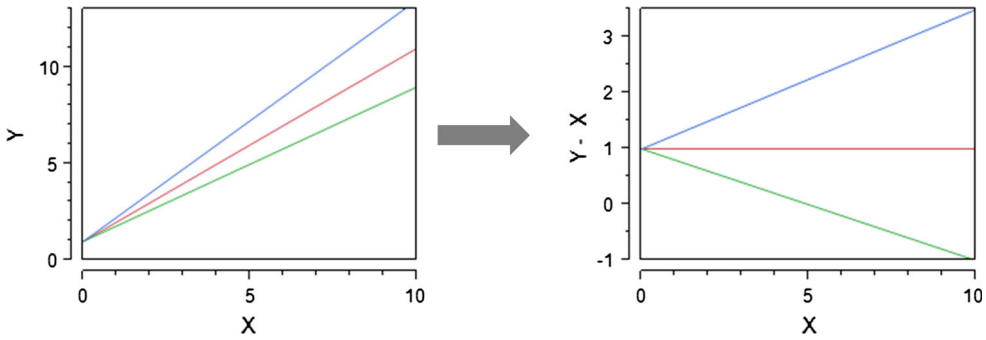

Fig. 4 Effect of offsetting transformation on three lines differing in slope with a common Y-intercept of +1. The (*red*) *line* with unit slope transforms to a line of zero slope and the *other two lines* transform to lines with slopes that reflect the extent to which the slope of the initial line deviates from unity

the 1-octanol/water partitioning system might conceivably be demonstrated (e.g. by relevant data analysis) to be optimal for creating LEMs, it cannot be assumed to be so simply on the basis that most drug discovery scientists do not look beyond it. The second problem arises when ionized forms of ligands bind to targets because it is usually only neutral forms of ionizable compounds that partition into the organic phase. When this is the case, neither (pIC_{50} − logP) nor (pIC_{50} − logD) serves to quantify the ease of transferring the bound form of the ligand from the organic phase to its binding site in the target. As such, when compounds bind to their targets in ionic forms it is incorrect to state,

> A molecule with an LLE equal to zero based on LogP, where target affinity is equal to LogP, can be thought of as having the same affinity for its target as it does for 1-octanol, whereas a drug candidate with an LLE of 6 has a one-million-fold higher affinity for its target compared to 1-octanol [1].

It should be noted that good correlations of activity with both ClogP and logD may be observed within structural series. One rationale for this is that ionizable groups, hydrogen bond donors and hydrogen bond acceptors are relatively conserved in these situations which are typically encountered when optimizing leads [2].

The acid/base properties of ligands [52] do appear to cause a certain degree of confusion with respect to definition of LEMs and it may be helpful to address this topic in more detail. The relationship between logD and logP can be conveniently written in terms of logP and F_{neut}, the fraction of ligand in its neutral form(s) [2].

$$\log D(pH) = \log P + \log F_{neut}(pH) \quad (14)$$

Equation 14 shows that logD is, in general, a function of pH and cannot exceed logP, equaling it only when functional groups that are significantly ionized at the relevant pH are absent from the molecular structure. It is logD that is actually measured and for ionizable compounds logP must be either derived from the logD-pH profile or calculated from logD using Eq. 14 when pK_a is known. One important question is which of logP or logD (at a specified pH) is the more appropriate measure of lipophilicity with which to offset potency. Equation 14 also shows that when logD is used to define an LEM we can make compounds more ligand efficient simply by increasing the extent to which they are ionized and drug discovery scientists might consider whether or not this would represent a prudent course of action. One scenario that is of particular interest is one in which ligands are predominantly ionized in physiological media and bind as ionized forms to their targets. This is likely to apply for amines and carboxylic acids when assayed at normal physiological pH. Although this is a common scenario, it is worth remembering that physiological pH is not always neutral (e.g. within lysosome) and pK_a values of functional groups such as sulfonamides [53] and guanidines [54] can, with appropriate substitution, be engineered to straddle neutral pH. Consider as a thought experiment an amine A for which logP = 3 and which is 90 % protonated under physiological conditions. Now consider amine B which is identical to A except that it is 99.9 % protonated. The amines share a common logP value of 3 but while A has a logD value of 2 that of B is 0. Since the amines are identical (apart from differing in pK_a) their cations bind with equal affinity and the pK_d values will differ only by 0.05 log units, reflecting the relative proportions (99.9/90 = 1.11) of the protonated forms of the two amines. This thought experiment suggests that, if the scenario in question applies, potency is likely to correlate more strongly with logP than with logD for predominantly ionized compounds. Analogous reasoning would suggest that logD would be the more appropriate lipophilicity descriptor for modelling phenomena like aqueous solubility and permeability for which the proportion of compound in the neutral form is usually limiting. However, the partitioning and ionization components of logD (Eq. 14) should ideally be decoupled for modelling so as to allow activity to respond differently to each factor.

A case study: ligand-efficiency-dependent lipophilicity

The use of molecular size and lipophilicity to define different LEMs raises the question of whether both properties can be

incorporated into a single LEM and LELP (ligand-efficiency-dependent lipophilicity) [45] has been defined as:

$$\mathrm{LELP} = (-\mathrm{HA} \times \mathrm{logP})/\Delta G^{\circ} \quad (15)$$

The LELP metric provides an instructive example with which to illustrate pitfalls which one can encounter when attempting to demonstrate value for an LEM. We examine it in some detail in the hope that it will encourage drug discovery scientists to frame their own questions when assessing the likely value of a metric. The functional form of LELP shows that it can have negative values and it can be described as an 'inverted' LEM in that its denominator is ΔG° and the risk factors are both in the numerator. One reason for defining LELP in this manner is that it avoids placing logP, which can be zero, in the denominator. Although LELP is at least defined for a zero value of logP, the metric does become completely insensitive to activity when this is the case. When LELP was introduced it was suggested that its value should lie between -10 and 10 for acceptable leads [45] and this implies an assumption that what might be termed 'lipophilicity risk' is symmetrical about a logP value of zero. An LEM like LELP that includes two risk factors needs to weight the contributions of each to the overall risk in an appropriate manner. For example, the risk associated with 25 heavy atoms and logP of 3 would be assessed using LELP as equal to those associated with 50 heavy atoms and logP of 1.5 or 75 heavy atoms and logP of 1. Others have made analogous observations about the LELP metric [28].

We focus on the TNK2012 study [55] which presents a case for using LELP in drug discovery projects. The study asserts that "multidimensional optimization toward oral drug candidates should deliver compounds with log P/log D between -1 and 3" [55] but fails to acknowledge that logP and logD are not in general identical. For example, the logD value for a carboxylic acid measured at neutral pH will typically be at least three units lower than the corresponding logP value. Even if we ignore the ambiguity in how the recommended lipophilicity range is specified, it is not symmetrical about zero and therefore not in accordance with LELP which considers lipophilicity risk to be symmetrical about a logP value of zero. There are obvious problems [28] with using LELP when the value of the chosen measure of lipophilicity approaches zero and the TNK2012 study states,

> Although LELP has its own limitation for compounds with logP < 1, due to the correlation of logP with molecular mass project compounds generally do not possess large heavy atom count with low log P [55].

The TNK2012 study fails to acknowledge that the problems associated with logP < 1 are not restricted to compounds with large heavy atom count and no evidence whatsoever is presented in support of the claimed correlation. The potential user of LELP can legitimately question the value of a metric that has a stated limitation across half the lipophilicity range recommended for oral drug candidates.

One of the aims of the TNK2012 study appears to be exploring the use of LEMs in combination although it makes no reference to $\mathrm{LLE_{AT}}$ [28] which is an obvious competitor as a metric in which activity is normalized by both molecular size and lipophilicity. If discrimination between different classes of compound (e.g. marketed drugs and leads) is to be the focus of the analysis, the simplest and most obvious way to do this is to plot one LEM against the risk factor for the other LEM. For example, we might plot ($\mathrm{pIC_{50}}$/HA) against ClogP or ($\mathrm{pIC_{50}} - \mathrm{ClogP}$) against HA and attempt to quantify the degree to which the descriptors in question separate the different classes of compound. The TNK2012 study eschews these data presentations in favor of a plot of ($\mathrm{pIC_{50}} - \mathrm{ClogP}$) against ClogP that displays less information (activity and lipophilicity but not molecular size). Another presentation of data used in the TNK2012 study is plot of LELP against LLE and this has been criticized [10] in the context of correlation inflation for its use of standard error to assert separation of compound classes. If separation of compound classes is actually the objective then one would need to show that the plot of LELP against LLE did this more effectively than alternatives such as a plot of ClogP against HA. Put another way, one would need to show that introduction of an additional parameter (ΔG°) increased the degree of separation of compound classes in order to validate the LEMs in this manner. The LELP and LLE functions share both activity and lipophilicity as arguments which introduces a correlation that affects the visual impact of the plot. When the plotted quantities are correlated, error bars will not, in general, be aligned with the axes and it is incorrect to simply assume that they are. The potential user of LELP can legitimately ask why plausible and potentially superior alternatives to the plot of LELP against LLE were not examined and how comparisons would have been made objectively had the alternatives been considered.

The TNK2012 study attempts to demonstrate the superiority of LELP over LLE by comparing relationships between each metric and a number of in vitro pharmacokinetic and safety parameters which are binned (high or low) for the analysis [55]. Mean values of each metric were presented with standard errors rather than standard deviations to indicate the spread in the measurements. This approach to presenting data has been criticized [10] as misleading because it exaggerates trends and makes relationships appear to become stronger as data sets become larger. When trends are weak, the fact that one might be stronger than the

other is overshadowed by the reality that each descriptor explains only a small proportion of the variance and neither is usefully predictive. The bigger question raised by this analysis is why LELP or LLE were even considered as predictors of in vitro pharmacokinetic and safety parameters. Why would one expect activity against arbitrary primary targets to influence these pharmacokinetic or safety parameters? If a compound shows activity against more than one target, which value of the LEM should be used? For these results to be meaningful, it would be necessary to demonstrate that LLE and LELP are more effective predictors of in vitro pharmacokinetic and safety parameters than ClogP and (HA × ClogP). Were this not the case, the effect of introducing of the additional parameter ($\Delta G°$) to the analysis would just be to simply add noise.

Recommendations and conclusions

Before making specific recommendations, it may be helpful to summarize what has been discussed so far. An LEM is a transformation of measured activity that is intended to quantify compound quality and can also be thought of as a simple QSAR model that relates activity to the physicochemical property(s) chosen to describe it. Defining an LEM requires that assumptions be made and the usefulness of an LEM is determined by the degree to which these assumptions hold. We have argued that there is no justification for making either of the assumptions (zero intercept or unit slope) that relate to scaling or offsetting measured activity and to do so distorts our perception of relationships between activity and physicochemical properties. Put another way, if we wish to quantify the extent to which a compound beats a trend, we need to first establish that trend and this means modelling the data, for example by using least squares regression to fit activity to risk factors such as molecular size and lipophilicity. The basis of least squares regression is to determine a function of the predictor variable(s) that minimizes the sum of the squares, $\Sigma(Y_{res})^2$ of the residuals, where the residual, Y_{res}, for a data point is defined as the difference between the value, Y_{obs}, observed for the dependent variable and the value, Y_{pred}, predicted for it by the model.

$$Y_{res} = Y_{obs} - Y_{pred} \quad (16)$$

The residual is also the most direct measure of the extent to which the activity of a compound beats the trend in the data.

There are a number of advantages to analyzing measured activity in this manner. Most importantly, it is the actual trend observed in the data that is used to normalize activity and Fig. 1 illustrates why this is important. This measure of the extent to which a compound's activity beats the trend is invariant with respect to the choice of standard or reference concentration and the uncertainty in residuals is independent of molecular size. Residuals have sign which makes it easy to see whether or not the activity of a compound beats the trend in the data. Modelling activity allows the effects of logP, ClogP, logD, molecular size and composite risk factors (e.g. HA × logP) to be explored within a common data-analytic framework and magnitudes of residuals can be viewed in the context of assay reproducibility. The effects of different risk factors can be analyzed simultaneously with multivariate models [56] and non-linear models can be used to explore the tendency of activity to plateau with molecular size without the assumptions about zero molecular size activity limits that distort analysis. Making residuals the focus of SARs also establishes a link between ligand efficiency and activity cliff concepts [57–59].

Modelling activity data yields more information than just residuals and parameters resulting from fitting activity to one or more risk factors can be informative. For example, a large positive intercept and relatively flat slope resulting from a linear fit of pIC_{50} to HA might be interpreted as binding that is dominated by a hot spot [60]. Parameters fit to data can be used to compare groups of compounds according to specific (e.g. structural series or scaffold) or generic (e.g. neutral at physiological pH) chemotype. When activity is strongly correlated with a risk factor, a steep response (slope) of activity to the risk factor is desirable. Measures of quality of fit such as the coefficient of determination (r^2) can also help project teams to assess alternative structural series. Provided that it is not due to assay noise or a narrow range in activity, a weak correlation between pIC_{50} and lipophilicity indicates that activity is driven by factors other than lipophilicity and that the project team has more room to maneuver.

It might be argued that a residual-based measure of the extent to which activity beats the trend in the data is not the absolute measure of compound quality that would be needed to decide whether or not to pursue a specific screening hit. To reject hits for which LipE falls below a cutoff is to make implicit assumptions about both the response of activity to lipophilicity and the profile (e.g. $IC_{50} < 10$ nM; ClogP < 3) required for a clinical candidate and it can be argued, in the interest of transparency, that assumptions like these should be articulated rather than hidden in a metric. Regardless of how projections are made from characteristics of hit compounds, it is still necessary to model activity data as the project evolves in order to check the validity of the underlying assumptions. There are also other factors which need to be accounted for when assessing the potential of a hit. For example, the availability of analogs could allow SAR to be explored without the need for synthesis. Alternatively, crystal structures may

suggest ways of linking two fragments in a strain-free manner with minimal steric footprint. The binding of a hit in a conformational, tautomeric or ionization state that is relatively high in energy should always be seen as an opportunity because synthetic modifications that stabilize the bound form will lead to an increase in affinity in the absence of unfavorable ligand–protein contacts [61].

It is important to remember that the extent to which properties such as lipophilicity and molecular size are predictive of in vivo behavior of compounds remains an issue even when the actual trend observed in the data is used to normalize activity. If we really want to understand the role of lipophilicity as a determinant of in vivo behavior, we will need to study relationships [48] between exposure of orally-dosed compounds and lipophilicity rather than add to an ever-growing list of metrics, guidelines, mnemonics and rules of thumb for drug-likeness, compound quality and 'developability'. Even when the observed global relationships between in vivo behavior and physicochemical properties are weak, the relevant correlations may prove to stronger within structural series. One possibility that should be anticipated is that lipophilicity and molecular size become more predictive of dose-normalized, free plasma levels as their values become more extreme. Analysis of exposure is likely to require non-linear models and data analysis methods [62] that allow in-range and out-of-range (e.g. concentration lower than detection limit) data to be analyzed together. Although the free level of a drug in plasma can be seen as a driver of its action, we should not lose sight of the fact that it is not currently possible to measure the free, intracellular concentration of an arbitrary compound in live humans.

So why should we consider ligand efficiency metrics to be harmful? Most importantly, as shown in Table 1 and Figs. 1, 2, 4 they distort our perception of the relationship between activity and the risk factor(s) with which we choose to normalize it. Neither the scaling nor the offsetting transformations used for normalizing activity has any real physiochemical basis and excessive reliance on metrics inhibits more thorough examination of data. We hope this Perspective will stimulate debate and encourage others to challenge the assumptions and opinions that shape and constrain drug discovery.

Acknowledgments We are grateful to the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, (Grants #2011/01893-3 and #2011/20572-3) and the Conselho Nacional de Pesquisa (CNPq) for financial support. We thank Peter Bernstein, Anna Linusson and the reviewers of the manuscript for their helpful and constructive advice.

References

1. Hopkins AL, Keserü GM, Leeson PD, Rees DC, Reynolds CH (2014) The role of ligand efficiency metrics in drug discovery. Nat Rev Drug Discov 13:105–121
2. Kenny PW, Montanari CA, Prokopczyk IM (2013) $ClogP_{alk}$: a method for predicting alkane/water partition coefficient. J Comput Aided Mol Des 27:389–402
3. Schultz MD (2013) Setting expectations in molecular optimizations: strengths and limitations of commonly used composite parameters. Bioorg Med Chem Lett 23:5980–5991
4. Schultz MD (2014) Improving the plausibility of success with inefficient metrics. ACS Med Chem Lett 5:2–5
5. Kenny PW, Montanari CA, Prokopczyk IM, Sala FA, Sartori GR (2013) Automated molecule editing in molecular design. J Comput Aided Mol Des 27:655–664
6. Kenny PW (2009) Hydrogen bonding, electrostatic potential and molecular design. J Chem Inf Model 49:1234–1244
7. Linusson A, Gottfries J, Lindgren F, Wold S (2000) Statistical molecular design of building blocks for combinatorial chemistry. J Med Chem 43:1320–1328
8. van de Waterbeemd H, Smith DA, Beaumont K, Walker DK (2001) Property-based design: optimization of drug absorption and pharmacokinetics. J Med Chem 44:1313–1333
9. Lipinski CA, Lombardo F, Dominy BW, 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
10. Kenny PW, Montanari CA (2013) Inflation of correlation in the pursuit of drug-likeness. J Comput Aided Mol Des 27:1–13
11. Muthas D, Boyer S, Hasselgren C (2013) A critical assessment of modeling safety-related drug attrition. Med Chem Commun 4:1058–1065
12. Erlanson DA, McDowell RS, O'Brien T (2004) Fragment-based drug discovery. J Med Chem 47:3463–3482
13. Leach AR, Hann MM, Burrows JN, Griffen EJ (2006) Fragment screening: an introduction. Mol BioSyst 2:429–446
14. Hajduk PJ (2006) Fragment-based drug design: how big is too big? J Med Chem 49:6972–6976
15. Albert JS, Blomberg N, Breeze AL, Brown AJH, Burrows JN, Edwards PD, Folmer RHA, Geschwindner S, Griffen EJ, Kenny PW, Nowak T, Olsson L, Sanganee H, Shapiro AB (2007) An integrated approach to fragment-based lead generation: philosophy, strategy and case studies from AstraZeneca's drug discovery programs. Curr Top Med Chem 7:1600–1629
16. Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) Recent developments in fragment-based drug discovery. J Med Chem 51:3661–3680
17. Blomberg N, Cosgrove DA, Kenny PW, Kolmodin K (2009) Design of compound libraries for fragment screening. J Comput Aided Mol Des 23:513–525
18. Erlanson DA (2012) Introduction to fragment-based drug discovery. Top Curr Chem 317:1–32
19. Joseph-McCarthy D, Campbell AJ, Kern G, Moustakas D (2014) Fragment-based lead discovery and design. J Chem Inf Model 54:693–704
20. Abad-Zapatero C, Perišić O, Wass J, Bento AP, Overington J, Al-Lazikani B, Johnson ME (2010) Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation. Drug Discov Today 15:805–811
21. Abad-Zapatero C, Champness EJ, Segall MD (2014) Alternative variables in drug discovery: promises and challenges. Future Med Chem 6:577–593
22. Smith DA, Di L, Kerns EH (2010) The effect of plasma protein binding on in vivo efficacy: misconceptions in drug discovery. Nat Rev Drug Discov 9:929–939
23. Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. Drug Discov Today 9:430–431
24. Tropsha A (2011) Best practices for QSAR model development, validation and exploitation. Mol Inform 29:476–488

25. Gilson MK, Given JA, Bush BL, McCammon JA (1997) The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophys J 72:1047–1069
26. Zhou H-X, Gilson MK (2009) Theory of free energy and entropy in noncovalent binding. Chem Rev 109:4092–4107
27. Gilson MK, Zhou H-X (2011) Calculation of protein-ligand binding affinities. Annu Rev Biophys Biomol Struct 36:21–42
28. Mortenson PN, Murray CW (2011) Assessing the lipophilicity of fragments and early hits. J Comput Aided Mol Des 25:663–667
29. Kuntz ID, Chen K, Sharp KA, Kollman PA (1999) The maximal affinity of ligands. Proc Natl Acad Sci 96:9997–10002
30. Smith RD, Engdahl AL, Dunbar JB, Carlson HA (2012) Biophysical limits of protein-ligand binding. J Chem Inf Model 52:2098–2106
31. Abraham MH (1993) Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes. Chem Soc Rev 22:73–83
32. Bissantz C, Kuhn B, Stahl M (2010) A medicinal chemist's guide to molecular interactions. J Med Chem 53:5061–5084
33. Hann MM, Leach AR, Harper G (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. J Chem Inf Comp Sci 41:856–864
34. Reynolds CH, Tounge BA, Bembenek SD (2008) Ligand binding efficiency: trends, physical basis, and implications. J Med Chem 51:2432–2438
35. Williams JW, Morrison JF (1979) The kinetics of reversible tight-binding inhibition. Methods Enymol 63:437–467
36. Abad-Zapatero C, Metz JT (2005) Ligand efficiency indices as guideposts for drug discovery. Drug Discov Today 10:465–469
37. Lloyd G, Czaplewskia LG, Collins I, Boyd EA, Brown D, East SP, Gardiner M, Fletcher R, Haydon DJ, Henstock V, Ingram P, Jones C, Noula C, Kennison L, Rockley C, Rose V, Thomaides-Brears HB, Ure R, Whittaker M, Stokes NR (2009) Antibacterial alkoxybenzamide inhibitors of the essential bacterial cell division protein FtsZ. Bioorg Med Chem Lett 19:524–527
38. Ladbury JE, Klebe G, Freire E (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. Nat Rev Drug Discov 9:23–27
39. Lewis ML, Cucurull-Sanchez L (2009) Structural pairwise comparisons of HLM stability of phenyl derivatives: introduction of the Pfizer metabolism index (PMI) and metabolism-lipophilicity efficiency (MLE). J Comput Aided Mol Des 23:97–103
40. Holdgate GA, Gill AL (2011) Kinetic efficiency: the missing metric for enhancing compound quality? Drug Discov Today 16:910–913
41. Murray CW, Erlanson DA, Hopkins AL, Keserü GM, Leeson PD, Rees DC, Reynolds CH, Richmond NJ (2014) Validity of ligand efficiency metrics. ACS Med Chem Lett ASAP. 10.1021/ml500146d
42. Verdonk ML, Rees DC (2008) Group efficiency: a guideline for hits-to-leads chemistry. ChemMedChem 3:1179–1180
43. Leeson PD, Springthorpe B (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. Nat Rev Drug Discov 6:881–890
44. Ryckmans T, Edwards MP, Horne VA, Correia AM, Owen DR, Thompson LR, Tran I, Tutt MF, Young T (2009) Rapid assessment of a novel series of selective CB2 agonists using parallel synthesis protocols: a lipophilic efficiency (LipE) analysis. Bioorg Med Chem Lett 19:4406–4409
45. Keserü GM, Makara GM (2009) The influence of lead discovery strategies on the properties of drug candidates. Nat Rev Drug Discov 8:203–212
46. Freeman-Cook KD, Hoffman RL, Johnson TW (2013) Lipophilic efficiency: the most important efficiency metric in medicinal chemistry. Future Med Chem 5:113–115
47. Mannhold R, Poda GI, Ostermann C, Tetko IV (2009) Calculation of molecular lipophilicity: state-of-the-art and comparison of log P methods on more than 96,000 compounds. J Pharm Sci 98:861–893
48. Valko K, Chiaparin E, Nunhuck S, Montanari D (2012) *In vitro* measurement of drug efficiency index to aid early lead optimization. J Pharm Sci 101:4155–4169
49. McTigue M, Murray BW, Jeffrey H. Chen JH, Denga Y-L, Solowiej J, Kania RS (2012) Molecular conformations, interactions, and properties associated with drug efficiency and clinical performance among VEGFR TK inhibitors. Proc Nat Acad Sci109:18281–18289
50. Schultz MD (2013) The thermodynamic basis for the use of lipophilic efficiency (LipE) in enthalpic optimizations. Bioorg Med Chem Lett 23:5992–6000
51. Abraham MH, Chadha HS, Whiting GS, Mitchell RC (1994) Hydrogen bonding. 32. An analysis of water-octanol and water-alkane partitioning and the ΔlogP parameter of Seiler. J Pharm Sci 83:1085–1100
52. Manallack DT, Prankerd RJ, Yuriev E, Oprea TI, Chalmers DK (2013) The significance of acid/base properties in drug discovery. Chem Soc Rev 42:485–496
53. Bell PH, Roblin RO (1942) Studies in chemotherapy. VII. A theory of the relation of structure to activity of sulfanilamide type compounds. J Am Chem Soc 64:2905–2917
54. Taylor PJ, Wait AR (1986) σ_i Values for heterocycles. J Chem Soc Perkin Trans 2:1765–1770
55. Tarcsay A, Nyiri K, Keserű GM (2012) Impact of lipophilic efficiency on compound quality. J Med Chem 55:1252–1260
56. Kalivas J (1999) Interrelationships of multivariate regression methods using eigenvector basis sets. J Chemom 13:111–132
57. Maggiora GM (2006) On outliers and activity cliffs—why QSAR often disappoints. J Chem Inf Model 46:1535
58. Guha R, Van Drie JH (2008) Structure-activity landscape index: identifying and quantifying activity cliffs. J Chem Inf Model 48:646–658
59. Stumpfe D, Bajorath J (2012) Exploring activity cliffs in medicinal chemistry. J Med Chem 55:2932–2942
60. DeLano WL (2002) Unraveling hot spots in binding interfaces: progress and challenges. Curr Opin Struct Biol 12:14–20
61. Black E, Breed J, Breeze AL, Embrey K, Garcia R, Gero TW, Godfrey L, Kenny PW, Morley AD, Minshull CA, Pannifer AD, Read J, Rees A, Russell DJ, Toader D, Tucker J (2005) Structure-based design of protein tyrosine phosphatase-1B inhibitors. Bioorg Med Chem Lett 15:2503–2507
62. Lind E (2010) QSAR analysis involving assay results which are only known to be greater than, or less than some cut-off limit. Mol Inform 29:845–852