



Pushing the Ligand Efficiency Metrics: Relative Group Contribution (RGC) Model as a Helpful Strategy to Promote a Fragment “Rescue” Effect

Andrés Felipe Vásquez^{1,2} and Andrés González Barrios^{1*}

¹ Grupo de Diseño de Productos y Procesos (GDPP), School of Chemical Engineering, Universidad de los Andes, Bogotá, Colombia, ² Laboratorio de Fisiología Molecular, Instituto Nacional de Salud, Bogotá, Colombia

OPEN ACCESS

Edited by:

Jose L. Medina-Franco,
National Autonomous University of
Mexico, Mexico

Reviewed by:

Sergio Hidalgo Figueroa,
Instituto Potosino de Investigación
Científica y Tecnológica (IPICYT),
Mexico
Salvatore Guccione,
University of Catania, Italy

*Correspondence:

Andrés González Barrios
andgonza@uniandes.edu.co

Specialty section:

This article was submitted to
Medicinal and Pharmaceutical
Chemistry,
a section of the journal
Frontiers in Chemistry

Received: 19 May 2019

Accepted: 24 July 2019

Published: 16 August 2019

Citation:

Vásquez AF and González Barrios A
(2019) Pushing the Ligand Efficiency
Metrics: Relative Group Contribution
(RGC) Model as a Helpful Strategy to
Promote a Fragment “Rescue” Effect.
Front. Chem. 7:564.
doi: 10.3389/fchem.2019.00564

The ligand efficiency (LE) indexes have long been used as decision-making criteria in drug discovery and development. However, in the context of fragment-based drug design (FBDD), these metrics often exhibit a strong emphasis toward the selection of highly efficient “core” fragments for potential optimization, which are not usually considered as parts of a larger molecule with a size typical for a drug. In this study, we present a relative group contribution (RGC) model intended to predict the efficiency of a drug-sized compound in terms of its component fragments. This model could be useful not only in rapidly predicting all the possible combinations of promising fragments from an earlier hit discovery stage, but also in enabling a relatively low-LE fragment to become part of a drug-sized compound as long as it is “rescued” by other high-LE fragments.

Keywords: ligand efficiency metrics, fragment-based screening, property-based design, drug discovery, protein-ligand interactions, structure-activity relationship, fragment library

INTRODUCTION

Ligand efficiency metrics have been applied as a decision-making strategy nearly universally accepted during the last two decades (Hopkins et al., 2014). They are intended to compare the quality of hits and leads during hit to lead and lead optimization phases of drug discovery (Cavalluzzi et al., 2017). Among these metrics, the ligand efficiency (LE), firstly described by Hopkins et al. (2004), continues to be the most widely used index to discriminate between promising molecules and those which are not (Reynolds, 2015). LE is defined by the following formula:

$$LE = \frac{\Delta G}{N} \quad (1)$$

where $\Delta G = -RT \ln K_d$, N (also known as HAC or heavy atom count) represents the number of heavy, non-hydrogen atoms, and K_d corresponds to the equilibrium dissociation constant.

As shown in Equation (1), LE normalizes the potency by size, specifically representing the average contribution ΔG (Gibbs free energy) per heavy atom. LE is typically used in FBDD as cut-off criterion to retrieve just high-LE fragments in a screening process (Murray and Verdonk, 2006; Schultes et al., 2010). Intriguingly, because LE usually considers fragments as independent chemical entities, some other metrics have emerged to consider the change in affinity as a fragment is developed into a larger, high-affinity drug-sized compound. One prime example is

the group efficiency (GE), described in 2008 (Verdonk and Rees, 2008), which allows measuring the contribution to the binding efficiency of a particular group of atoms added to an existing lead molecule:

$$GE = \frac{\Delta\Delta G}{\Delta N} \quad (2)$$

where $\Delta\Delta G = \Delta G(B) - \Delta G(A)$ and $\Delta N = (N(B) - N(A))$; in other words, $\Delta\Delta G$ is equal to the difference between the Gibbs free energy of the existing molecule (or fragment) “A” and the new combined molecule “B” and ΔN corresponds to the difference between the number of non-hydrogen atoms of molecules “A” and “B.” However, GE is based on a pairwise comparison of structurally closely related compounds and, hence, it is frequently applied for optimization of a high-efficient fragment (Hopkins et al., 2014). Therefore, a rapid and simple method for comparing efficiency of different fragments as part of a whole, including if they are dissimilar to each other (or if they occupy different pockets in the target molecule), needs to be developed.

Several independent studies in the past two decades have indicated an overemphasis on potency by the pharmaceutical industry (Albert et al., 2007; Hopkins et al., 2014). Still, other factors such as chemical novelty (Medina-Franco et al., 2014), selectivity fine-tuning (Costantino and Barlocco, 2018), structural alerts avoidance (Jasial et al., 2017), and synthetic accessibility (Fukunishi et al., 2014) are increasingly playing a key role in drug design. Considering these non-mutually exclusive events, we hypothesize that fragments that not necessarily exhibit a high efficiency level during a screening procedure, either virtual or experimental, would still have the possibility of taking part in a complete drug-sized compound.

In this study, we propose that a relative group contribution (RGC) model based on the efficiency of its component fragments may estimate the efficiency of a drug-sized compound. This model calculates the minimum efficiency required for unknown fragments by considering the efficiency of those already known, which facilitates a rapid elucidation of the best combinations of fragments. Likewise, this model facilitates that fragments with a relatively low efficiency may not necessarily be eliminated at an early stage of the screening process and, consequently, may become eventually represented as chemical moieties within the final candidate compound -a phenomenon herein referred to as fragment “rescue” effect.

THEORETICAL FRAMEWORK OF THE RGC MODEL

The proposed model is based on three main assumptions:

1. The efficiency of an entire molecule may be estimated as the weighted root mean square (WMRS) of the efficiency of its component fragments. The rationale for using this type of mean is intended to (1) cope with the negative values of ΔG and (2) consider effectively the potentially different number of non-hydrogen atoms (N) for each component fragment.
2. The efficiency of each fragment (regarded as the entire ratio and not just the quotient) is, in principle, dependent on each

other, excepting in cases of two fragments when the N for them is equal to each other (and then their weight is equivalent), or in cases when three or more fragments are involved.

3. The efficiency of each fragment is directly calculated from the ΔG resulting in its direct interaction with a specific location (i.e., binding site or pocket) in a particular receptor.

Our hypothesis assumes, according to its first principle, that the WRMS of the LEs of the fragments composing an entire molecule (\overline{LE}_q) corresponds to the actual LE of this latter (LE_T) (A comprehensive list of mathematical terms is shown in **Supplementary Material**). Because this mean is intended to be proportionally similar to the real, total LE of an entire molecule (LE_T), we refer herein to it as the apparent total LE (LE_T^{app}):

$$\overline{LE}_q = LE_T^{app} \approx LE_T \quad (3)$$

For clarity of the RGC concept, we consider first the LE_T^{app} as a simple arithmetic mean:

$$LE_T^{app} = \frac{1}{x}(LE_1 + LE_2 + LE_3 + \dots + LE_x) \quad (4)$$

where LE_i corresponds to the LE of the component fragments (LE_1 , LE_2 , etc.), and x refers to the number of fragments composing the molecule. Therefore, once LE is expressed in terms of the Equation (1):

$$LE_T^{app} = \frac{1}{x} \left(\frac{\Delta G_1}{N_1} + \frac{\Delta G_2}{N_2} + \frac{\Delta G_3}{N_3} + \dots + \frac{\Delta G_x}{N_x} \right) \cong \frac{\Delta G_T}{xN_T} \quad (5)$$

where ΔG_i is the change in Gibbs free energy for each composing fragment (ΔG_1 , ΔG_2 , etc.) up to a maximum number of fragments x and N_i correspond to the number of non-hydrogen atoms of each fragment. Similarly, ΔG_T and N_T represent the change in Gibbs free energy and number of non-hydrogen atoms for the entire molecule, respectively. Should be remembered that (5) is based on an arithmetic mean and hence it assumes that N is equal among all composing fragments, so that it would just be applicable in this specific scenario.

If we consider the Equation (5) for a molecule composed by a single fragment:

$$LE_T^{app} = \frac{1}{x} \sum_{i=1}^x \left(\frac{\Delta G}{N} \right)_i = LE_i \quad (6)$$

we can observe that LE_T^{app} correspond to the LE value of the unique component fragment, namely LE_1 , which supports a scenario where the component fragment is also the entire “final” molecule. However, expressing (5) for a molecule composed by two fragments:

$$LE_T^{app} = \frac{1}{x} \sum_{i=1}^x \left(\frac{\Delta G}{N} \right)_i = \frac{1}{2} \sum_{i=1}^2 \left(\frac{\Delta G}{N} \right)_i = \frac{1}{2}(LE_1 + LE_2) \quad (7)$$

we could notice that, in contrast to the one-fragment case, the existence of more than one compound allows for the solution of the equation in terms of a particular fragment:

$$LE_2 = 2LE_T^{app} - LE_1 \quad (8)$$

The last equation poses a simple but important principle: Starting from an “ideal” LE_T^{app} , a particular low-LE fragment can be successfully chosen or “rescued” by one or more high-LE fragments. Now, if we consider a three-fragment case:

$$LE_T^{app} = \frac{1}{x} \sum_{i=1}^x \left(\frac{\Delta G}{N} \right)_i = \frac{1}{3} \sum_{i=1}^3 \left(\frac{\Delta G}{N} \right)_i = \frac{1}{3} (LE_1 + LE_2 + LE_3) \quad (9)$$

Interestingly, for this case, even if we assume in this example that we know LE_1 , it is still possible to consider a LE value for the unknown fragments LE_2 and LE_3 grouping them together into a single term:

$$LE_T^{app} = \frac{1}{3} (LE_1 + 2LE_\delta) \quad (10)$$

where the LE delta (LE_δ) corresponds to a transient, “ideal” value intended to be equal for all the fragments which individual LE is still unknown. Therefore, as we will discuss below, this value will be modified as long as new LE values are known for fragments, independently of their position.

Likewise, assuming also that we just know LE_1 , we could express LE_T^{app} for the two-fragment case:

$$LE_T^{app} = \frac{1}{2} (LE_1 + LE_\delta) \quad (11)$$

which would indicate that:

$$LE_\delta = LE_2 \quad (12)$$

This result suggests, as we also discuss below, that a LE_δ is expected to equal the LE value of a last fragment to be known (LE_u), independently of the number of fragments and their position. This behavior corresponds to a subtractive average (SA). Remarkably, although the nature of LE_δ is somewhat similar to the cumulative average (CA) or moving average (MA), the number of fragments with unknown LE is continuously decreasing and LE_δ does not “run” within a predetermined window size.

Finally, assuming also that we just know LE_1 , we could express LE_T^{app} for the one-fragment case:

$$LE_T^{app} = [LE_1 + 0 (LE_\delta)] = LE_1 \quad (13)$$

indicating that LE_δ can only be calculated if there are at least two starting fragments and, more importantly, it is especially useful in cases of three or more of them.

At this point, think of the LE_T^{app} as a whole for an undetermined series of fragments:

$$LE_T^{app} = \frac{1}{x} (LE_1 + \dots + LE_x) \quad (14)$$

If we assume again that we just know LE_1 , it is possible to rearrange LE_T^{app} using LE_δ :

$$LE_T^{app} = \frac{1}{x} [LE_1 + (x-1) LE_\delta] \quad (15)$$

where $(x-1)$ corresponds to the coefficient of the LE_δ value independently of the number of starting fragments as shown in Equations (10, 11, 13). Hence, if we resolve for LE_δ :

$$LE_\delta = \frac{1}{x-1} (xLE_T^{app} - LE_1) \quad (16)$$

Now, if we take into account that the number 1 in this equation actually corresponds to the number of known LE values of fragments or a , we can observe that:

$$\text{if } (x-a) \rightarrow 1, \text{ then } LE_\delta \rightarrow LE_u \quad (17)$$

What means that the more $(x-a)$ tends to 1, the more LE_δ tends to LE_u , just as we saw previously in Equation (12). Finally, considering the formula for LE_δ in terms of an undetermined number of fragments with different known and unknown LE values:

$$LE_\delta = \frac{1}{x-a} \left[\sum_{i=1}^x \left(\frac{\Delta G}{N} \right)_i - \sum_{j=0}^a \left(\frac{\Delta G}{N} \right)_j \right] \quad (18)$$

$$\text{if and only if } \begin{cases} 1 \leq x < \infty \\ x \in \mathbb{N} \\ 0 \leq a < x \\ a \in \mathbb{Z}_0^+ \end{cases}$$

where the first summation term indicates the “ideal” sum of LE values for the existing fragments (LE_i) as if they would have the same value, and the second summation term refers to the “real” sum of all fragments which LE value is already known (LE_j). On the other hand, if $a = 0$, LE_0 would not proceed as a real value (and by extension the second summation term). Therefore, in this specific case a consequence would be that:

$$LE_\delta = LE_T^{app} \quad (19)$$

Now, after having explained the basic concepts of RGC and LE_δ , we could express LE_T^{app} in terms of the WMRS according to our hypothesis:

$$LE_T^{app} = \frac{\sum_{i=1}^x LE_i^2 w_i / \sum_{i=1}^x w_i}{\sum_{i=1}^x \left(\frac{\Delta G}{N} \right)_i^2 w_i / \sum_{i=1}^x w_i} \quad (20)$$

where LE_i corresponds to the LE of each component fragment, w_i refers to the weight of each fragment (depending on the N of each one) and x refers to the number of fragments.

Likewise, our formula for LE_δ would be:

$$LE_\delta = \frac{1}{\sqrt{w_\delta(x-a)}} \left[\sqrt{\sum_{i=1}^x \left(\frac{\Delta G}{N} \right)_i^2 w_i} - \sum_{j=0}^a \left(\frac{\Delta G}{N} \right)_j^2 w_j \right] \quad (21)$$

$$\text{if and only if } \begin{cases} 1 \leq x < \infty \\ x \in \mathbb{N} \\ 0 \leq a < x \\ a \in \mathbb{Z}_0^+ \end{cases}$$

The additional term in this equation, namely W_{δ} , corresponds to the “ideal” weight of all fragments with unknown LE value, as if they would have the same value. Just as with LE_{δ} , this parameter is expected to change with every new LE value of fragment known, until the value (weight) corresponding to the last fragment with unknown LE value is adopted.

FRAGMENT SELECTION AND LE_T PREDICTION BY THE RGC MODEL

As presented in a hypothetical example (Figure 1), three central premises may be elucidated for the RGC model by selecting hit compounds, starting from a cut-off LE value:

1. A low-efficiency fragment (i.e., with a $LE < LE_T^{app}$) can be rescued IF there exists high-efficiency fragments (i.e., with a $LE > LE_T^{app}$) in ALL the other positions (i.e., binding sites or pockets) of the target molecule, usually a protein. If this condition is not satisfied, the fragment could be automatically rejected from the set of possible combinations of fragments for an entire molecule.
2. If there are no high-efficiency fragments in all the other positions simultaneously, a molecular fragment with low-efficiency fragment can still be rescued IF at least one fragment in any other position with an efficiency high enough to reach exists, in average with the first, the LE_T^{app} . This implies, therefore, that once a fragment is rescued by one or more high-efficiency fragments in other positions, any fragment in subsequent unexplored positions is just required to have a mid-efficiency (with a $LE \cong LE_{\delta}$ ideal) as a minimum. Additionally, each time LE_{δ} changes, a differential “pushing” over the LE of unknown fragments occurs in terms of ΔG and/or N , which could be potentially modified to achieve an acceptable LE value and could therefore be selected.
3. Even if a low-efficiency fragment is not rescued after implementing the strategies stated in Equation (1, 2), a potential rescue could still take place exploring a more diverse library sample of fragments, assuming that (1) you are dealing with a number of fragments well-below under the maximum theoretical chemical space for them (a population of about 10^7 molecules) and (2) the fragment is not an outlier compared to other fragments intended to combine.

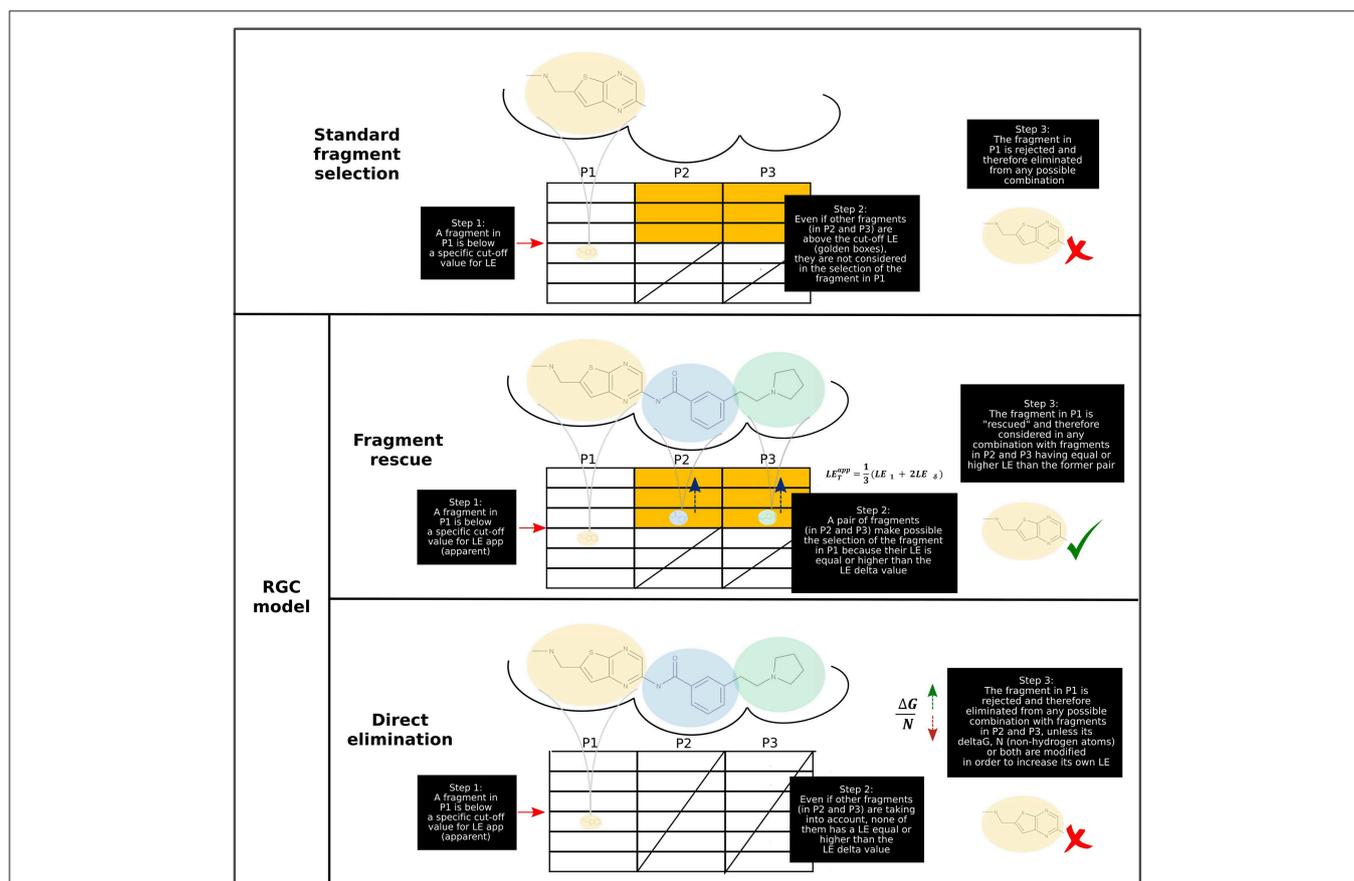


FIGURE 1 | Application of RGC model to a hypothetical fragment-based drug design (FBDD) campaign. **(Upper)** In the “standard” or classical screening approach, a fragment is selected (i.e., can be part of a final drug-size compound) depending exclusively upon their own LE. If this parameter is not equal or greater than a pre-established cut-off value, the fragment is rejected. **(Lower)** According to the RGC model, a fragment is selected depending on the fragments on the other positions. Based on the presence of high-LE fragments in alternative positions (illustrated by yellow boxes), a low-LE fragment may become either rescued or rapidly discarded (using the dynamic LE_{δ} value in both cases).

In addition, and according to our preliminary results, we found that LE_T^{app} values predicted by this model were consistent with the LE_T values calculated experimentally from a set of 16 drug-sized molecules taken from scientific literature (Figure 2, Table 1S).

DISCUSSION

The present study pretends to propose the RGC model as an innovative and effective approach to apply in drug design. This model and, especially, the fragment “rescue” effect that is conceptually implicit, offer an alternative for the long-standing FBDD paradigm of designing compounds merely based on the intrinsic binding energy of fragments, facilitating the introduction of other decision-making criteria that are becoming increasingly common.

If the principles of the RGC model are considered together, it is possible to elucidate two major advantages. First, we count on a limited amount of data and in order to more clearly reveal any trends, LE_T^{app} appears to increase as much as the LE_T , and there appears to be no dramatic shift toward higher efficiencies for particular fragments or protein targets. Secondly, since a particular fragment could be directly rejected early in the process and there are many fragments by pocket in a

typical FBDD campaign, this model might dramatically reduce the computational and synthetic costs, respectively (which is especially true in cases of three or more pockets).

The RGC model is, however, not free of inherent shortcomings. As a LE-derived metric, all fragments are assumed to maintain equal orientations both individually and as part of a larger chemical compound (Zartler and Shapiro, 2008), and phenomena such as hot spots (Zerbe et al., 2012; Rathi et al., 2017) or synergy (also called “super-additivity”) (Hebeisen et al., 2008; Nazaré et al., 2012) are not directly considered. Likewise, because its average-based nature, ΔG of each fragment is normalized not only at the number of non-hydrogen atoms but also on the number of component fragments. Therefore, the less accurate (or more extreme) ΔG values for fragments in each position are, the greater the difference expected between the LE_T^{app} and the LE_T . However, we believe that the impact of these hurdles could be minimized, improving the confidence of any potential fragment “rescue,” if additional energy terms derived from rigid body barrier (ΔG_{rigid}), linker binding ($\Delta G_{binding}$) or strain (ΔG_{strain}) (Murray and Verdonk, 2006; Cherry and Mitchell, 2008) are included and the ΔG values are both accurate and above a reasonable cut-off value.

A final examination about the implications of this work leads us to assert that the fragment “rescue” phenomenon is far from being new: it has already occurred and continues to

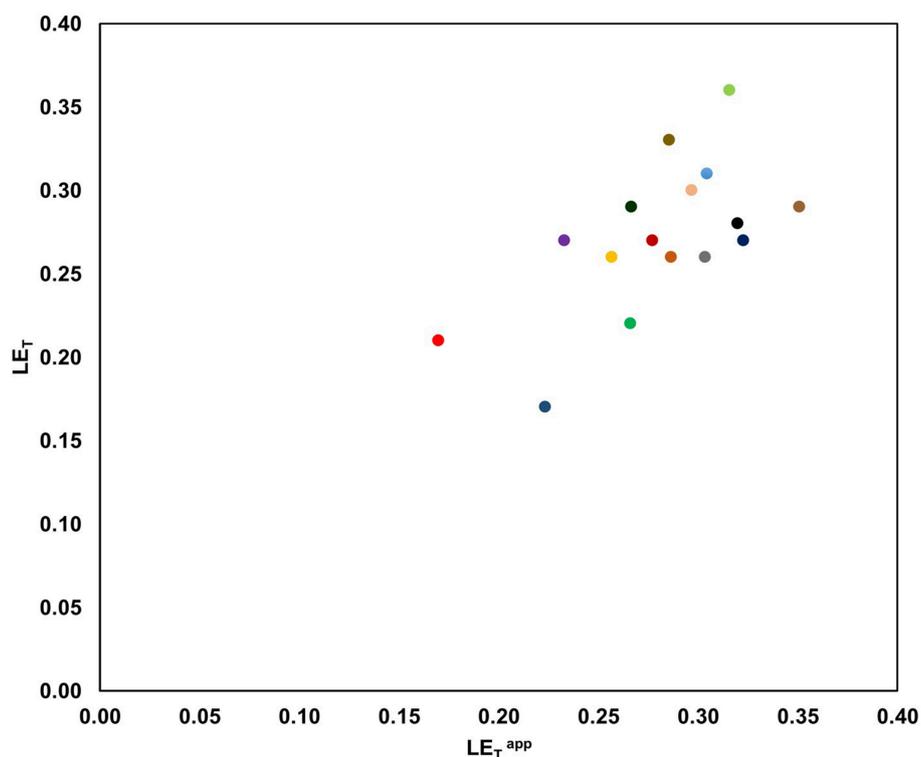


FIGURE 2 | Preliminary comparison between LE_T^{app} and LE_T for a set of 16 drug-sized literature compounds developed during FBDD studies. All compounds were elaborated using a linking strategy (on two fragments) for 10 different protein targets [●/● LDHA ● Replication Protein A (RPA) ● Blood coagulation factor Xa ● Bcl-2 ● DOT1L ● Hsp90 ● Pantothenate synthetase (PtS) ● Blood coagulation factor XIa ● CK2 ● BACE1 ● Endothiapepsin (Epn) ● Bcl-x_L ● PKM2].

occur, but usually during a lead optimization instead an earlier hit discovery phase. The rationale for this statement lies in two central facts. First, we currently know that the LE value of a compound tends to decrease during optimization process (Bembenek et al., 2009), while both its lipophilicity and its MW (and, hence, the number of non-hydrogen atoms -N) tends to increase (Ferenczy and Keseru, 2016). Second, the energy of supramolecular interactions is widely known to be largely different depending on the chemical moiety involved and, thus, ionic and hydrogen bonds are expected to account for a larger part of the drug-receptor binding energy (i.e., its ΔG) compared with hydrophobic interactions (Ermondi and Caron, 2006). Therefore, it is decidedly inviting to believe that, in many drug discovery initiatives, both the increase in LE and the decrease in MW and lipophilicity observed during lead optimization-could be explained by the addition of large and hydrophobic chemical moieties such as those cyclic aliphatic, which have a much smaller $\Delta G/N$ ratio compared to other fragments. Interestingly, these aliphatic moieties have been recently suggested by some authors to be more “developable” compared its aromatic analogs, which could be an additional factor behind this phenomenon (Lovering et al., 2009).

The RGC model presented in this study is based on the assumption that the LE of a drug-sized molecule may be estimated using the relative contribution of each component fragment. We believe this model could serve as a complementary benchmark for medicinal chemists in experimental or virtual fragment-based screening campaigns. Likewise, we consider that the RGC model could be implemented with other metrics based on either LE or a potency/size ratio and could be eventually adjusted to consider not only “linking” but also “growing” or “merging” as alternative fragment elaboration strategies.

REFERENCES

- Albert, J. S., Blomberg, N., Breeze, A., Brown, A., Burrows, J., Edwards, P., et al. (2007). An integrated approach to fragment-based lead generation: philosophy, strategy and case studies from AstraZeneca's Drug Discovery Programmes. *Curr. Top. Med. Chem.* 7, 1600–1629. doi: 10.2174/156802607782341091
- Bembenek, S. D., Tounge, B. A., and Reynolds, C. H. (2009). Ligand efficiency and fragment-based drug discovery. *Drug Discov. Today* 14, 278–283. doi: 10.1016/j.drudis.2008.11.007
- Cavalluzzi, M. M., Mangiatordi, G. F., Nicolotti, O., and Lentini, G. (2017). Ligand efficiency metrics in drug discovery: the pros and cons from a practical perspective. *Expert Opin. Drug Discov.* 12, 1087–1104. doi: 10.1080/17460441.2017.1365056
- Cherry, M., and Mitchell, T. (2008). “Introduction to fragment-based drug discovery,” in *Fragment-Based Drug Discovery*, eds E. R. Zartler and M. J. Shapiro, 1–32. doi: 10.1002/9780470721551.ch1
- Costantino, L., and Barlocco, D. (2018). “Designing Approaches to Multitarget Drugs,” in *Drug Selectivity: An Evolving Concept in Medicinal Chemistry*, 161–205.
- Ermondi, G., and Caron, G. (2006). Recognition forces in ligand-protein complexes: blending information from different sources. *Biochem. Pharmacol.* 72, 1633–1645. doi: 10.1016/j.bcp.2006.05.022
- Ferenczy, G. G., and Keseru, G. M. (2016). “Ligand efficiency metrics and their use in fragment optimizations,” in *Fragment-Based Drug Discovery Lessons and*

DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

AV planned and performed the entire *in silico* work presented in this study, contributed to the analysis and interpretation of data, and assisted the writing, editing, and submission of this manuscript. AG made substantial contributions to the data analysis, critical revision for important intellectual content, and document editing. All authors have read and approved the final manuscript.

FUNDING

We acknowledge funds from the Colombian Department of Science, Technology and Innovation COLCIENCIAS Grant No. 727.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Professor Marco de Vivo from Istituto Italiano di Tecnologia (IIT), in Genoa (Italy) for helpful discussions and recommendations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2019.00564/full#supplementary-material>

Outlook, eds D. A. Erlanson and W. Jahnke (Wiley-VCH Verlag GmbH & Co. KGaA), 75–98. doi: 10.1002/9783527683604.ch04

- Fukunishi, Y., Kurosawa, T., Mikami, Y., and Nakamura, H. (2014). Prediction of synthetic accessibility based on commercially available compound databases. *J. Chem. Inf. Model.* 54, 3259–3267. doi: 10.1021/ci500568d
- Hebeisen, P., Kuhn, B., Kohler, P., Gubler, M., Huber, W., Kitas, E., et al. (2008). Allosteric FBPase inhibitors gain 10(5) times in potency when simultaneously binding two neighboring AMP sites. *Bioorganic Med. Chem. Lett.* 18, 4708–4712. doi: 10.1016/j.bmcl.2008.06.103
- Hopkins, A. L., Groom, C. R., and Alex, A. (2004). Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* 9, 430–431. doi: 10.1016/S1359-6446(04)03069-7
- Hopkins, A. L., Keseru, G. M., Leeson, P. D., Rees, D. C., and Reynolds, C. H. (2014). The role of ligand efficiency metrics in drug discovery. *Nat. Rev. Drug Discov.* 13, 105–121. doi: 10.1038/nrd4163
- Jasial, S., Hu, Y., and Bajorath, J. (2017). How frequently are pan-assay interference compounds active? Large-scale analysis of screening data reveals diverse activity profiles, low global hit frequency, and many consistently inactive compounds. *J. Med. Chem.* 60, 3879–3886. doi: 10.1021/acs.jmedchem.7b00154
- Lovering, F., Bikker, J., and Humblet, C. (2009). Escape from flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* 52, 6752–6756. doi: 10.1021/jm901241e
- Medina-Franco, J. L., Martinez-Mayorga, K., and Meurice, N. (2014). Balancing novelty with confined chemical space in modern drug discovery. *Expert Opin. Drug Discov.* 9, 151–165. doi: 10.1517/17460441.2014.872624

- Murray, C. W., and Verdonk, M. L. (2006). "Entropic consequences of linking ligands," in *Fragment-Based Approaches in Drug Discovery*, eds R. Mannhold, H. Kubinyi, G. Folkers, W. Jahnke, and D. A. Erlanson (Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA), 55–66. doi: 10.1002/3527608761.ch3
- Nazaré, M., Matter, H., Will, D. W., Wagner, M., Urmann, M., Czech, J., et al. (2012). Fragment deconstruction of small, potent factor xa inhibitors: exploring the superadditivity energetics of fragment linking in protein-ligand complexes. *Angew. Chemie Int. Ed.* 51, 905–911. doi: 10.1002/anie.201107091
- Rathi, P. C., Ludlow, R. F., Hall, R. J., Murray, C. W., Mortenson, P. N., and Verdonk, M. L. (2017). Predicting "Hot" and "Warm" Spots for Fragment Binding. *J. Med. Chem.* 60, 4036–4046. doi: 10.1021/acs.jmedchem.7b00366
- Reynolds, C. H. (2015). Ligand efficiency metrics: why all the fuss? *Future Med. Chem.* 7, 1363–1365. doi: 10.4155/fmc.15.70
- Schultes, S., De Graaf, C., Haaksma, E. E. J., De Esch, I. J. P., Leurs, R., and Krämer, O. (2010). Ligand efficiency as a guide in fragment hit selection and optimization. *Drug Discov. Today Technol.* 7, 157–162. doi: 10.1016/j.ddtec.2010.11.003
- Verdonk, M. L., and Rees, D. C. (2008). Group efficiency: a guideline for hits-to-leads chemistry. *Chem. Med. Chem.* 3, 1179–1180. doi: 10.1002/cmdc.200800132
- Zartler, E. R., and Shapiro, M. J. (eds.). (2008). "Designing a fragment process to fit your needs," in *Fragment-Based Drug Discovery* (John Wiley & Sons, Ltd.), 15–37. doi: 10.1002/9780470721551.ch2
- Zerbe, B. S., Hall, D. R., Vajda, S., Whitty, A., and Kozakov, D. (2012). Relationship between hot spot residues and ligand binding hot spots in protein-protein interfaces. *J. Chem. Inf. Model.* 52, 2236–2244. doi: 10.1021/ci300175u

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Vásquez and González Barrios. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.