DOI: 10.1002/cmdc.200800132

Group Efficiency: A Guideline for Hits-to-Leads Chemistry

Marcel L. Verdonk* and David C. Rees*[a]

Herein we describe the concept of group efficiency (GE), which is an extension of ligand efficiency (LE), that we find particularly useful during the hit-to-lead and optimisation stages of drug discovery projects. LE has already become popular amongst medicinal chemists.[1–3] It is used to normalise the binding affinity of a compound with respect to its molecular weight and is a simple way to rank and compare the affinities of compounds with different sizes. The term LE was first suggested by Hopkins et al.[4] as a measure of the free energy of binding (ΔG_b) divided by the molecular size and is related to a publication from Kuntz et al.:[5]

$$\mathrm{LE} = -\Delta G_b/\mathrm{HAC}\ (\mathrm{units} = \mathrm{kcal\,mol^{-1}\ per\ heavy\ atom}) \quad (1)$$

for which $\Delta G_b = RT\ln K_d$, or approximated as $\Delta G_b \approx RT\ln \mathrm{IC}_{50}$, and HAC (heavy atom count) is the number of non-hydrogen atoms in the molecule. Table 1 lists some simple examples of ΔG_b and LE values for various molecular weights (M_r) and potencies. On average a molecule of $M_r = 500$ contains ~36 non-hydrogen atoms. Hence, if a binding affinity of 10 nM is required for a drug candidate, a compound of 500 Da needs to have LE ≈ 0.31 kcal mol^{-1} per atom and a 10 nM compound with $M_r = 300$ would have LE ≈ 0.52. When starting lead optimisation we favour a relatively high LE because this allows atoms to be added to modulate in vivo properties while still ending up with a candidate with a molecular weight that fits the Lipinski guidelines. However, the maximum possible LE varies according to the target.[6] For example, in our experience, it has been possible to obtain leads with high LEs (LE = 0.4–0.5) against kinase targets, whereas the LEs of published protein–protein interaction inhibitors have been estimated at ~0.24.[7] The LE concept has been extended to other physical properties of a ligand, such as its lipophilicity. This field was recently reviewed by Abad-Zapatero and Metz.[8]

Table 1. Examples of ΔG_b values for various potencies, and LE as a function of M_r and potency.

Potency [K_d]	ΔG_b	LE [kcal mol^{-1} per heavy atom]				
		$M_r = 100$ (HAC≈7)	$M_r = 200$ (HAC≈14)	$M_r = 300$ (HAC≈22)	$M_r = 400$ (HAC≈29)	$M_r = 500$ (HAC≈36)
1 mM	−4.12	0.59	0.29	0.19	0.14	0.11
1 μM	−8.24	1.18	0.59	0.37	0.28	0.23
1 nM	−12.35	1.76	0.88	0.56	0.43	0.34

Ligand efficiency is a property of an entire molecule. It is additionally useful to estimate the binding efficiency of parts of a molecule, or of groups added to an existing lead. This we refer to as group efficiency (GE). GE represents the binding efficiency of a functional group that has been added to an existing molecule "A" to form molecule "B", and is defined in a completely analogous manner to LE as the change in binding energy divided by the change in the number of non-hydrogen atoms:

$$\mathrm{GE} = -\Delta\Delta G_b/\Delta\mathrm{HAC} \quad (2)$$

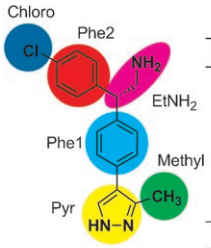
for which $\Delta\Delta G_b = \Delta G_b(\mathrm{B}) - \Delta G_b(\mathrm{A})$ and $\Delta\mathrm{HAC} = \mathrm{HAC(B)} - \mathrm{HAC(A)}$; in other words, the GE of the added functional group is its contribution to the free energy of binding per heavy atom.

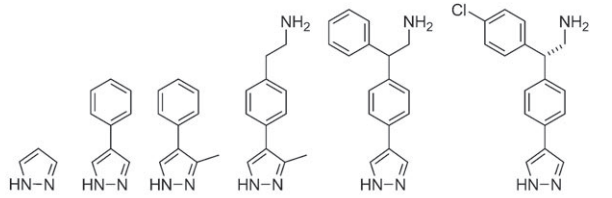
Given the relevant structure–affinity relationships (SAR), the GEs of the various parts of a lead compound can be derived. An example is given in Figure 1 for the kinase PKB (Akt).[1] To derive the GEs of the various parts of the molecule, we first applied a "Free–Wilson" analysis[9] on the SAR for this compound series, which provides the $\Delta\Delta G_b$ values for different groups; these are then simply divided by the ΔHACs to yield the GEs. Figure 1 shows that most of the affinity is provided by the pyrazole "hinge binder" and that the GE of the pyrazole is very high (GE = 1.5). It also highlights the fact that the 4-chloro substituent is a very efficient addition to the molecule (GE = 1.6). A similar analysis of the efficiency of various parts of a cofactor bound to ketopantoate reductase was reported by Ciulli et al.[10]

GE is a more sensitive metric to define the quality of an added group than a comparison of the LE of the parent and newly formed compounds. Assume, for example, that a phenyl ring is added to a parent compound "A" that contains 25 atoms, and that a 10-fold improvement in potency is obtained for compound "B". The LEs of the two compounds are very similar: 0.33 for A and 0.31 for B, and based on the LE values alone it would be tempting to conclude that the phenyl has been a decent addition to molecule A. However, the GE of the added phenyl group is only 0.23, which indicates that in fact it has been a relatively poor addition in comparison with the rest of the molecule.

A useful extension of the GE concept is that it provides simple guidelines for how much gain in potency should be aimed for as a function of the size of the added functional group. Table 2 lists these guidelines for three different GEs. For the above example in which a phenyl ring is added to a molecule, a gain in potency of at least 22-fold

[a] M. L. Verdonk, D. C. Rees
Astex Therapeutics Ltd., 436 Cambridge Science Park
Milton Road, Cambridge CB4 0QA (UK)
E-mail: m.verdonk@astex-therapeutics.com
d.rees@astex-therapeutics.com





pyrazole 1 2 3 4 5

Compound	Pyr	Phe1	Me	$EtNH_2$	Phe2	Cl	ΔG_b
pyrazole	x						–3.1[a]
1	x	x					–5.7
2	x	x	x				–6.0
3	x	x	x	x			–7.6
4	x	x		x	x		–9.0
5	x	x		x	x	x	–10.6
$\Delta\Delta G_b$	–7.3[b]	–2.5	–0.32	–1.6	–1.7	–1.6	
GE	1.5	0.42	0.32	0.54	0.28	1.6	

[a] Pyrazole inhibited PKB too weakly to determine an IC_{50} value in the bioassay; the value presented is an estimate based on crystallographic assessment of changes in the occupancy of pyrazole (in PKA-PKB) with varying compound concentration.
[b] Calculated as $\Delta\Delta G_b = \Delta G_b - \Delta G_{rigid}$, using ΔG_{rigid} = 4.2 kcal mol^{-1} (see Saxty et al.[1]).

Figure 1. Example of how a "Free–Wilson" analysis can be used to derive the GE values for various parts of a lead series for PKB.[1] The GEs are colour-coded according to functional group.

Table 2. Potency gain to be aimed for as a function of the number of added atoms.

ΔHAC	Fold improvement GE = 0.31 (M_r = 500)	GE = 0.39 (M_r = 400)	GE = 0.52 (M_r = 300)
1	1.7	1.9	2.3
2	2.8	3.6	5.5
3	4.6	6.8	13
4	7.7	13	30
5	13	24	71
6	22	46	170
7	36	88	390
8	60	170	920
9	100	320	2200
10	170	600	5100
11	280	1100	12000
12	460	2200	28000

[a] Data are shown for GE = 0.31 (assuming the project works towards a lead of 500 Da), GE = 0.39 (working towards a lead of 400 Da), and GE = 0.52 (when a lead of 300 Da is the aim).

should be obtained, just in order to stay on track for a 10 nM compound with M_r = 500. If lower molecular weight leads are the aim, even more potency should be gained (46-fold for GE = 0.39). GEs also make it very straightforward to compare the efficiencies of added groups of different sizes; for example, an indole substituent that provides a 20-fold increase in potency has GE = 0.19, whereas an imidazole substituent that provides a 15-fold potency jump has GE = 0.32.

There are a number of caveats for GE, two of which are mentioned here. Firstly, the effects of the added group are assumed to be independent of other groups within the molecule, and this is often an oversimplification. Effects such as conformational or electronic changes induced by the added group are included in its GE, even though the induced changes in interactions may occur distal to this group. Frequently in our laboratories, X-ray crystallographic data are available and can be used to help determine if this is an issue. Secondly, GE specifically addresses binding affinity and the size of the ligand. It should be used with caution for any SAR other than affinity and clearly it is of much less use when optimising in vivo activity.

In conclusion, in our laboratories we find GE a useful extension to the LE concept which, by referring to the guidelines presented in Table 2, can be easily applied by medicinal chemists in hit-to-lead and fragment-based lead discovery projects.

[1] G. Saxty, S. J. Woodhead, V. Berdini, T. G. Davies, M. L. Verdonk, P. G. Wyatt, R. G. Boyle, D. Barford, R. Downham, M. D. Garrett, R. A. Carr, *J. Med. Chem.* **2007**, *50*, 2293–2296.
[2] M. J. Palmer, A. S. Bell, D. N. Fox, D. G. Brown, *Curr. Top. Med. Chem.* **2007**, *7*, 405–419.
[3] P. M. Pelphrey, V. M. Popov, T. M. Joska, J. M. Beierlein, E. S. Bolstad, Y. A. Fillingham, D. L. Wright, A. C. Anderson, *J. Med. Chem.* **2007**, *50*, 940–950.
[4] A. L. Hopkins, C. R. Groom, A. Alex, *Drug Discovery Today* **2004**, *9*, 430–431.
[5] I. D. Kuntz, K. Chen, K. A. Sharp, P. A. Kollman, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9997–10002.
[6] P. J. Hajduk, *J. Med. Chem.* **2006**, *49*, 6972–6976.
[7] J. A. Wells, C. L. McClendon, *Nature* **2007**, *450*, 1001–1009.
[8] C. Abad-Zapatero, J. T. Metz, *Drug Discovery Today* **2005**, *10*, 464–469.
[9] S. M. Free, J. W. Wilson, *J. Med. Chem.* **1964**, *7*, 395–399.
[10] A. Ciulli, G. Williams, A. G. Smith, T. L. Blundell, C. Abell, *J. Med. Chem.* **2006**, *49*, 4992–5000.

Received: April 30, 2008
Published online on July 23, 2008

