Ligand Binding Efficiency: Trends, Physical Basis, and Implications

Charles H. Reynolds,*,† Brett A. Tounge,† and Scott D. Bembenek‡

Johnson & Johnson Pharmaceutical Research and Development, L.L.C., Welsh and McKean Roads, Spring House, Pennsylvania 19477, and 3210 Merryfield Road, San Diego, California 92121

Received October 4, 2007

Ligand efficiency (i.e., potency/size) has emerged as an important metric in drug discovery. In general, smaller, more efficient ligands are believed to have improved prospects for good drug properties (e.g., bioavailability). Our analysis of thousands of ligands across a variety of targets shows that ligand efficiency is dependent on ligand size with smaller ligands having greater efficiencies, on average, than larger ligands. We propose two primary causes for this size dependence: the inevitable reduction in the quality of fit between ligand and receptor as the ligand becomes larger and more complex and the reduction in accessible ligand surface area on a per atom basis as size increases. These results have far-ranging implications for analysis of high-throughput screening hits, fragment-based approaches to drug discovery, and even computational models of potency.

Introduction

A certain amount of additivity is implicit in most medicinal chemistry programs where different regions of a lead molecule are often optimized separately, in hopes of finding the best combination of moieties for a prospective drug. Further, efforts to improve potency almost inevitably lead to an increase in molecular size since additional van der Waals (hydrophobic) contacts between ligand and protein are generally favorable. This trend toward larger ligands in the pursuit of potency has been the topic of many papers and is generally considered detrimental to drug properties such as bioavailability.[1] Concerns about optimizing potency while keeping molecular size in check have led to the development of a simple metric that provides an estimate of the binding free energy per structural element, e.g., ligand binding efficiency. Ligand efficiency[2–4] is most commonly defined as the ratio of the free energy of binding over the number of heavy atoms in a molecule. Of course, the reciprocal, number of heavy atoms per log unit of activity, is an equally effective and in some ways more intuitive measure of efficiency.

Andrews et al.[4] attempted to exploit the presumed additivity of functional groups by developing a group additivity scheme for ligand binding. In their approach, common organic functional groups were assigned a group contribution according to the average contribution of that group to potency across a range of ligands. The goal was to develop a method for computing average binding affinities for any putative ligand based on its constituents. This computed average activity could then be compared to the experimental activity in order to assess whether the compound in question was an average, above-average, or below-average binder. This was followed by a paper from the Kuntz group[5] that examined the maximal affinities observed for any ligand against any target across a range of molecular sizes. Their analysis showed that maximal affinities (the best binders) varied greatly according to the type of protein–ligand interaction, with covalent and metal–ligands being the strongest. They also observed that the maximal ligand efficiency was not linear with ligand size, but seemed to decline as ligands became larger.

We have re-examined the concept of maximal ligand efficiency by studying a large number of protein–ligand complexes. Since our intent was to focus on classical small-molecule drug discovery, we have omitted the metals and small ions that were such efficient binders in the previous study from consideration. We have also taken advantage of the wealth of data now available. Our first objective was largely empirical and focused on evaluating how ligand efficiency varies across protein targets and ligand size. This was the subject of an earlier paper,[6] where we found a strong dependence of ligand efficiency as a function of ligand size. Our second objective, and the main focus of this study, was to unravel some of the underlying mechanisms responsible for this behavior.

Procedure

The protein targets in this study were chosen to provide a variety of active sites in terms of size, mechanism, and overall hydrophobicity/hydrophilicity. Similarly, the ligands were chosen to represent a range of chemotypes and, in particular, a wide range of molecular sizes. We assembled a large database of ligands and protein targets using the BindingDB database developed at the University of Maryland Biotechnology Institute.[7] We extracted 2581 ligands with reported K_i values and 6072 ligands with IC_{50} values. The final set covers 28 distinct protein targets, has K_i values that range from 0.001 nM to 213 μM. and IC_{50} values that range from 0.01 nM to 5.5 mM. The number of heavy atoms in both data sets covers a broad range of 6–78 and 7–62, respectively, for K_i and IC_{50}. To compute the ligand efficiencies, first the number of non-hydrogen heavy atoms (HA)[a] was calculated for each ligand. This number was then combined with the pK_i or pIC_{50} to calculate ligand efficiencies using the equation affinity/HA, where the affinity was approximated using either the pK_i or pIC_{50}. It is also possible to convert the K_i values (or less rigorously the IC_{50} values) to free energies first in order to compute ΔG /HA. In

* To whom correspondence should be addressed. Tel: 215-628-5675. Fax: 215-628-4985. E-mail: creynol1@prdus.jnj.com.
† Spring House, PA.
‡ San Diego, CA.

[a] HA, number of heavy atoms; FQ, fit quality; LE, ligand efficiency; LE_Scale, scaling factor to convert ligand efficiency to fit quality; SAR, structure–activity relationship; SASA, solvent-accessible surface area.

10.1021/jm701255b CCC: $40.75
Published on Web 04/02/2008

Table 1. Protein Targets for K_i and IC_{50} Data Sets

target	data set[a]	target	data set[a]
AchE	K_i/IC_{50}	MAO-B	K_i/IC_{50}
CYP19	K_i/IC_{50}	plasmepsin-II	K_i
CA-I	K_i	thrombin	K_i
CA-II	K_i/IC_{50}	BuChE	IC_{50}
CA-IV	K_i/IC_{50}	PKA	IC_{50}
Caspase-1	K_i/IC_{50}	FGFR-1	IC_{50}
Caspase-3	K_i/IC_{50}	GSK-3IC_{50}	
CDK-2	K_i/IC_{50}	HIV-1 RT	IC_{50}
CDK-4	K_i/IC_{50}	neuraminidase-A	IC_{50}
factor-Xa	K_i/IC_{50}	neuraminidase-B	IC_{50}
HIV-1 protease	K_i	PTP1B	IC_{50}
MMP-1	K_i	VEGFr-2	IC_{50}
MMP-9	K_i	farnesyl transferase	$_{50}$
MAO-A	K_i/IC_{50}	DPP-IV	K_i/IC_{50}

[a] Present in K_i, IC_{50}, or both data sets.

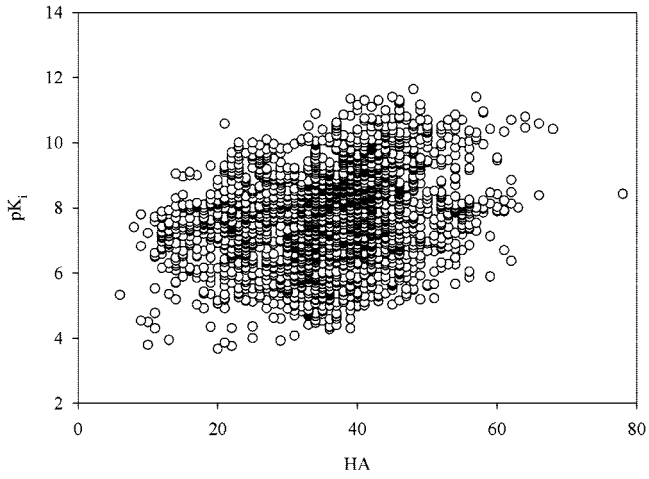

Figure 1. Plot of pK_i versus number of heavy atoms.

practical terms, the only difference is that the ΔG/HA efficiencies are systematically larger (i.e., $\Delta G = 1.37pK_i$).

We modeled the ligand structures using Maestro[8] and associated tools. The structures and energies were determined using the OPLS force field[9] and the GB/SA water solvation model.[10,11] The conformational searches were performed using a Monte Carlo torsional search algorithm.[12] In each case, a minimum of 2000 Monte Carlo steps were employed. The molecular properties, such as solvent accessible surface area, were calculated using QikProp.[13] All statistical analysis used either our internal 3DX program or Sigmaplot.[14]

Results and Discussion

Analysis of Binding Data. Due to the fundamental differences in the K_i and IC_{50} measurements[15] (K_i is directly related to free energy while IC_{50} values are relative), these two data types were analyzed separately. The protein targets included in each set are summarized in Table 1, and plots of the affinities versus size are given in Figures 1 and 2 for the pK_i and pIC_{50} sets, respectively. Affinities can vary widely for any given heavy atom count since it is possible to have ligands of any size that bind poorly because they simply do not have the correct geometry or functionality to bind to the active site. But if one considers the best (i.e., most potent) ligands at each size, a trend emerges. For example, in the K_i set, the change in the maximum binding affinity (pK_i) as a function of heavy atom count is initially fairly linear and quite steep (Figure 1), but the increase in potency lags as the number of heavy atoms increases. A similar trend is observed for the pIC_{50} data (Figure 2). In order to make this trend clearer, Figure 3 is a plot of only the most

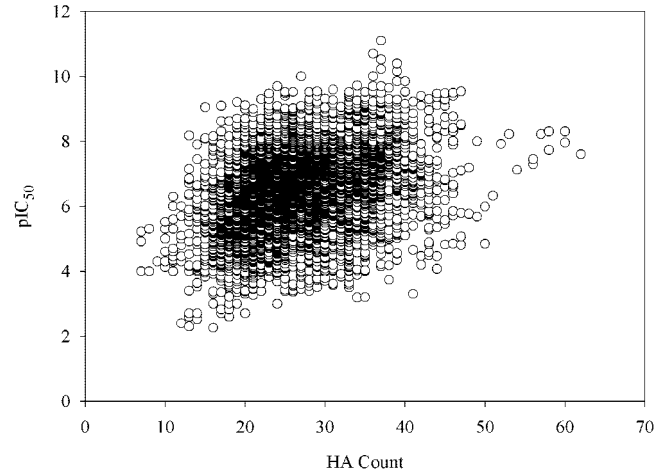

Figure 2. Plot of pIC_{50} versus number of heavy atoms.

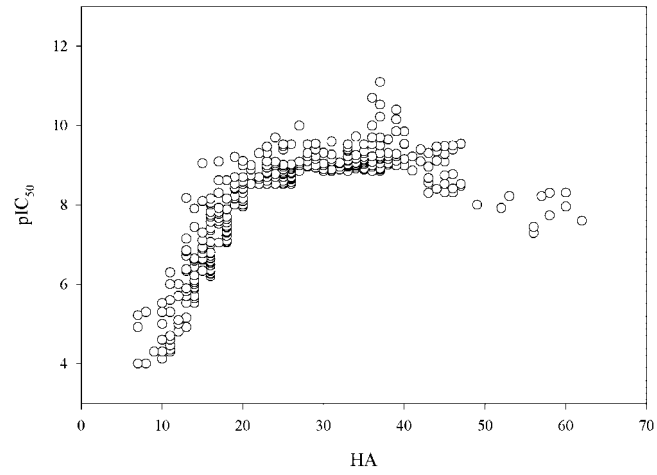

Figure 3. Plot of only the most potent inhibitors at each size. The "maximal affinities" as measured by pIC_{50} increase rapidly up to 20 heavy atoms, but plateau beyond 25.

potent ligands in each size regime across all targets for the pIC_{50} data. There is a marked flattening of the potencies as one reaches approximately 20 heavy atoms. Another way to illustrate this phenomenon is to plot the ligand efficiencies (e.g., pK_i/HA) versus molecular size. Figures 4 and 5 show pK_i and pIC_{50} based ligand efficiencies plotted against the number of heavy atoms. If potency were linearly related to the number of atoms in a ligand this plot should be clustered around a constant value and should not change with size. Clearly, this is not the case. In both plots, a very dramatic decline is observed in ligand efficiency as size increases. This drop is most precipitous in the early part of the curve, e.g., between 10 and 20 heavy atoms, and flattens toward very large sizes (e.g., above 40 heavy atoms). Given the number of ligands and targets we considered, this would seem to be a rather general trend.

While the trends are relatively consistent across all protein targets for both the K_i and IC_{50} sets, there are some differences. As one might expect, the curve can be shifted slightly to the right or left depending on the properties of the protein target. Large open targets such as HIV protease and Plasmepsin-II have a ligand distribution that is biased toward higher heavy atom counts. By comparison, carbonic anyhdrase-II and acetylcholinesterase have smaller active sites, and their ligand distributions are shifted toward smaller compounds (Figure 6). Summary statistics for the K_i and IC_{50} data sets are given in Table 2.

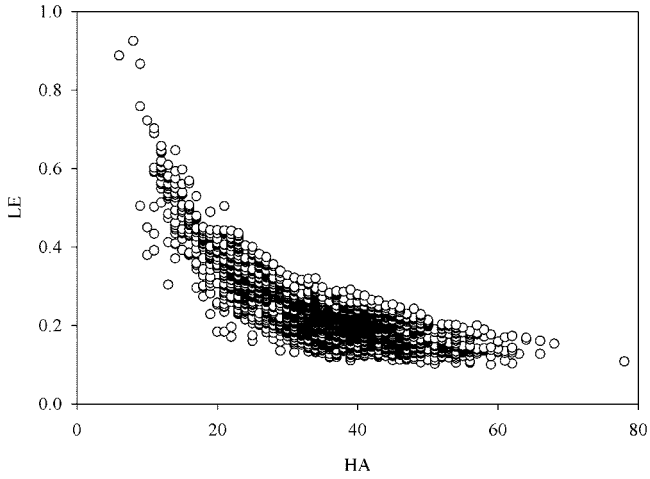

Figure 4. Ligand efficiency (LE = pK_i/HA) as a function of number of heavy atoms for the pK_i data set. Ligand efficiency falls off dramatically between 10 and 25 heavy atoms.

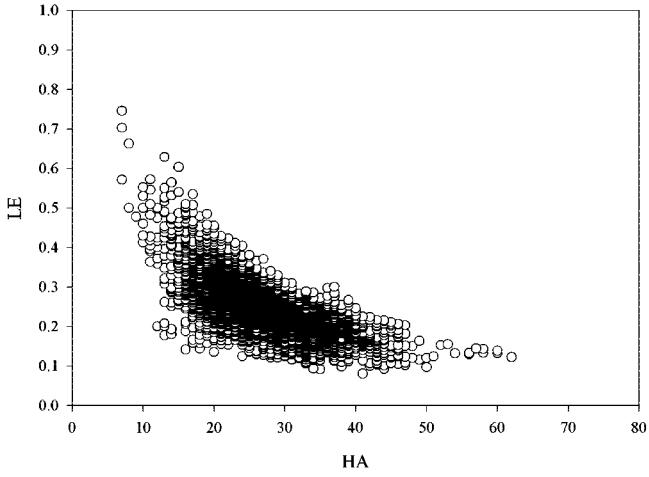

Figure 5. Ligand efficiency (LE = pIC_{50}/HA) as a function of heavy atoms for the IC_{50} data set. Ligand efficiency shows a similar precipitous decline between 10 and 25 heavy atoms.

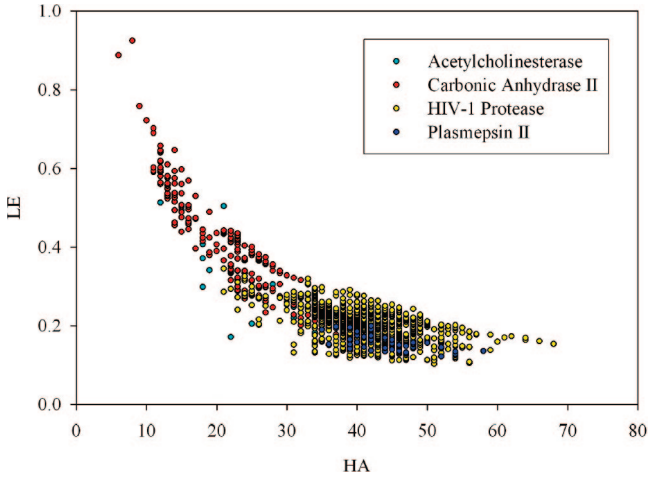

Figure 6. LE plotted verses the heavy atom count (HA) for different subsets of the K_i data set.

Fit Quality Score. One of the primary consequences of the trends discussed above is that ligand efficiency cannot be evaluated independently of ligand size. Stated differently, small ligands have inherently greater average ligand efficiency than large ligands and any comparison between small and large compounds must take this difference into account. In previous work, we suggested an alternative score for ligand "fit quality" that was normalized so that the most efficient binders in the IC_{50} data set were scaled to have a score of 1.0 across a wide range of molecular sizes. Since the K_i data should be more reliable, we refit the ligand efficiency scaling using this set. We also expanded the fit range to cover ligands with 10–50 heavy atoms. The new fit quality score (eqs 1 and 2) is given by

$$\text{LE\_Scale}=0.0715+7.5328/(\text{HA})+25.7079/(\text{HA}^2)-361.4722/(\text{HA}^3) \quad (1)$$

$$\text{FQ}=\text{LE}/\text{LE\_Scale} \quad (2)$$

where HA is the number heavy atoms, FQ is the fit quality, and LE is the ligand efficiency. A plot of the fit and the resulting FQ scores for the K_i data set is shown in Figure 7. The size-normalized fit quality score makes it easy to identify ligands with good ligand efficiencies or even ligands with exceptional efficiencies (e.g., fit quality scores greater than 1.0). Examination of the IC_{50} data set (Figure 8) illustrates this point. The compounds with the best ligand efficiency for a given atom count are easily identified as those that fall around 1.0 in terms of their fit quality score. Those with particularly high ligand efficiency given their atom count fall well above 1.0. A few representative structures for the carbonic anhydrase K_i set are given in Figure 9. Structure **a** in Figure 9 has a very high ligand efficiency and an FQ scrore of 1. Structure **b** is about twice as large in terms of heavy atoms and has a lower ligand efficiency but the same FQ score of 1. This FQ score indicates that although the raw ligand efficiency is lower, this ligand provides optimal binding for its size. Structure **c** is the same size as **b**, but has even poorer ligand efficiency relative to **b** and an FQ score consistent with much less optimal binding.

As described previously, the scaling parameter LE_Scale (Table 3) was derived by fitting ligand efficiency values for ligands with 10 to 50 heavy atoms. Given the small number of compounds under 15 heavy atoms and the predominance of carbonic anyhdrase inhibitors in this size range, we suggest using

Table 2. Summary Statistics

database		min	max	avg	std dev
K_i					
	HA	6	78	36	10
	pK_i	3.7	11.7	7.7	1.4
	LE	0.10	0.92	0.24	0.09
	SASA	265	1348	785	173
IC_{50}					
	HA	7	60	26	6.5
	pIC_{50}	2.3	11.1	6.6	1.3
	LE	0.08	0.75	0.26	0.07
	SASA	231	1295	1146	264

Table 3. Selected Values for Scaling Parameter (LE_Scale)

heavy atoms	pK_i	ΔG	heavy atoms	pK_i	ΔG
10	0.7204	0.9855	32	0.3210	0.4391
12	0.6686	0.9145	34	0.3061	0.4187
14	0.6090	0.8331	36	0.2928	0.4006
16	0.5545	0.7586	38	0.2809	0.3843
18	0.5074	0.6941	40	0.2702	0.3696
20	0.4672	0.6391	42	0.2605	0.3564
22	0.4331	0.5925	44	0.2517	0.3443
24	0.4039	0.5525	46	0.2437	0.3334
26	0.3787	0.5181	48	0.2363	0.3233
28	0.3569	0.4882	50	0.2295	0.3140
30	0.3378	0.4621			

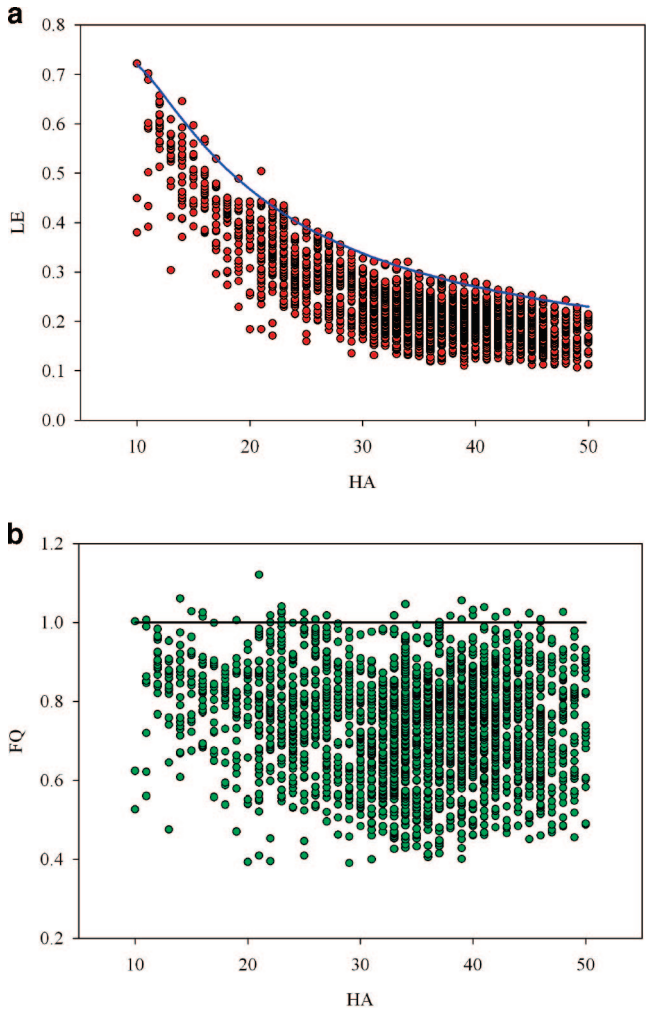

Figure 7. Original ligand efficiency as a function of heavy atoms for the K_i data set is shown in the red circles. These values were scaled using the fit represented by the blue line (a) to produce the fit quality metric (green) shown in (b). Fit quality scores around 1 (black line in (b)) indicated a near-optimal ligand binding affinity for a given number of heavy atoms.

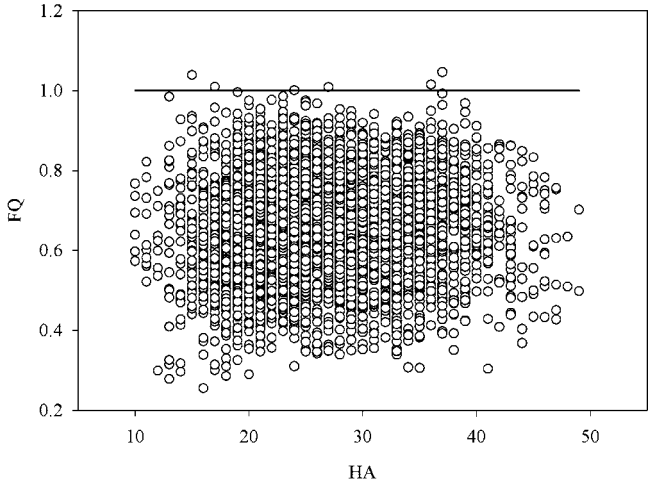

Figure 8. Fit quality (FQ) values for the IC_{50} data set.

the LE_Scale value at 15 for any compound with 15 or fewer heavy atoms. Similarly, we would use the LE_Scale value at 50 for any compounds with more than 50 heavy atoms. Indeed, for drug discovery applications one might consider limiting the

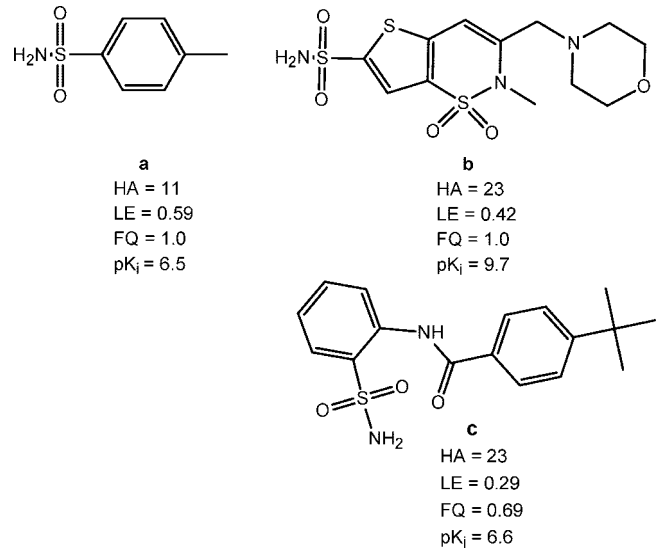

Figure 9. Examples of high and low FQ scoring ligands from the carbonic anhydrase K_i set.

scaling value at a lower heavy atom threshold in order to penalize larger ligands. These limits avoid the extreme values that might arise for very small or very large compounds. This scheme should be particularly valuable in the context of fragment-based drug discovery where the objective is to identify small fragments that bind to the protein of interest and either combine fragments or grow from a fragment to develop potent ligands. Ligand efficiency is often applied to fragments and the analysis above shows that small fragments (e.g., 15 or fewer heavy atoms) should be assessed differently than larger structures in terms of raw ligand efficiencies. Beyond fragment-based design, we see the fit quality score as a very useful metric in all phases of drug discovery. It provides a direct assessment of the quality of binding (or efficiency) that can be used to compare ligands of any size. As such the fit quality score gives added guidance during all phases of an SAR program as to whether atoms are being budgeted wisely with regard to their impact on potency. The scaling parameters (LE_Scale) can also be useful for scaling other computational results that are often confounded with molecular size, such as raw docking scores.

Physical Interpretation. We hypothesize that a variety of factors might lead to the observed reduction in ligand binding efficiency as molecular size increases. The first factor is largely enthalpic and relates to the fact that structural compromises are inevitable when any ligand binds to a protein. As ligands, and their corresponding active sites, become larger the number of accommodations might also be expected to increase as the ligand is required to satisfy more constraints. One might also expect entropy to play a significant role. Indeed Andrews et al. explicitly included a crude correction factor in their additivity scheme that they ascribed to loss of rotational entropy in the bound ligand. A last factor is simply geometric and relates to the fact that the amount of surface area available to interact with a protein active site is smaller on a per atom basis for larger ligands.

Structural Constraints. Large complex ligands have many points of contact with a protein active site that must be satisfied for potent binding. Satisfying multiple molecular recognition sites all at one time with a single ligand inevitably leads to structural compromises that reduce affinity. This is analogous to the concept of ligand complexity put forward by Hann and co-workers[16] and has also been alluded to by Murray and

Table 4. Calculated ΔE Values and Hydrogen Bond Distances for the Model Active Site

	ΔE (kcal/mol)	acid distance[a]	amine distance[a]
fragments	−17.9	1.591	1.557
tether 4	−9.6	3.314	1.581
tether 5	−15.9	1.707	1.720
tether 6	−16.4	1.586	1.548
tether 7	−13.9	1.602	1.561

[a] Hydrogen bond distance (angstroms) between the acid or amine in the fragment/ligand and the model Lys or Asp.

Verdonk in their analysis of small fragment binding.[17] Simply put, in the case of a small fragment that interacts with a single site in the protein the fragment is free to adopt a geometry that is ideal to make the most energetically favorable interaction possible (e.g., a salt bridge between a small carboxylic acid and an arginine in the active site). If, however, the ligand is larger and must span two protein interaction sites it becomes more difficult to achieve an optimal fit with both functional groups in both binding pockets. As the size and complexity increase further to three or even more specific protein–ligand interactions, structural compromises in the form of induced strain and suboptimal protein–ligand complementarity multiply. This can be demonstrated with a trivial example. We carried out a series of calculations for a model system consisting of an isolated lysine and aspartic acid separated in space by 15 Å. First we allowed two fragments, methylamine and acetic acid, to bind separately to this model active site. In each case, the ligand fragments formed ideal salt bridge pairs with their corresponding partners, i.e., methylamine with the Asp and acetic acid with the Lys. We then tethered the acid and amine fragments together with successively larger alkyl chains and reoptimized the structure (Table 4). This constraint due to the tether prevents both model interaction sites from being optimally accommodated at the same time. Indeed in the case of a short tether (e.g., four methylenes) only one of the two interactions can be maintained and the total energy relative to the two fragments is very high (Table 4). The minimum energy complex results from a tether containing six intervening methylenes. Even so, this complex is approximately 1.5 kcal/mol less stable than the independent fragments. This difference is due to small perturbations in the two acid–base salt bridges that are necessitated by the tether. It must be remembered that chemical structures are not infinitely tunable. They are assembled using discrete distances and angles between atoms that are enforced by nature. For example, a C−C single bond is always going to be approximately 1.5 Å even if the distance really needed to achieve an optimal fit is 1.2 or 1.8. As the tether is increased beyond the optimal length of six methylenes, the ligand is capable of interacting with both sites in our crude active site, but the ligand is forced to fold and incur an internal strain penalty. Examples of tethered and independent fragments bound to this model system are shown in Figure 10. The interaction energies and hydrogen bond pair distances are summarized in Table 4. Of course, any real receptor or enzyme target is much more complex than this simple model with many more independent interaction points and geometric constraints that must be accommodated. Thus it is not surprising that larger and more complicated ligands might bind less efficiently due to the attenuating effects of greater ligand strain and suboptimal ligand-protein interactions.

Ligand Surface Area. One of the most straightforward reasons for the observed size dependence of ligand efficiencies is related to the accessible surface area presented by ligands of different number of atoms. To a first approximation, the magnitude of protein–ligand interactions, particularly for hydrophobic interactions, is limited by the surface area on a ligand that is available to interact with the target protein. To investigate how this available surface area changes as a function of the number of heavy atoms, we calculated the solvent accessible surface area (SASA) for a model system of alkanes as well as for our K_i and IC_{50} data sets. These SASA values are shown as a function of the heavy atom count in Figure 11. As expected, the SASA term grows as the atom count increases for the alkanes, but the gain per unit atom trails off. For example at 10 heavy atoms the SASA/atom is 46 compared to 36 for 40 heavy

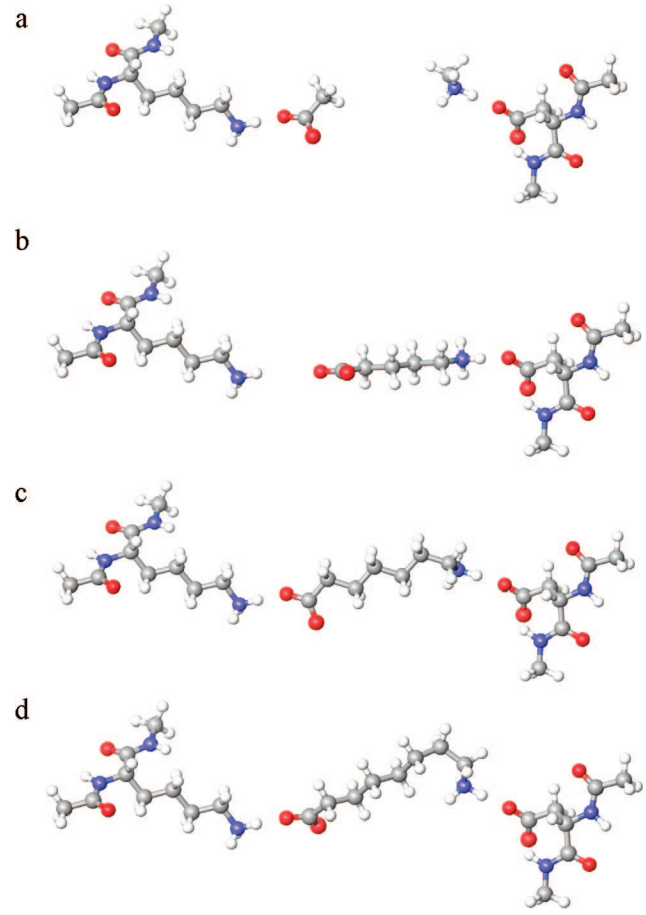

Figure 10. Optimized structure for independent fragments (a), four methylene tethers (b), six methylene tethers (c), and seven methylene tethers (d).

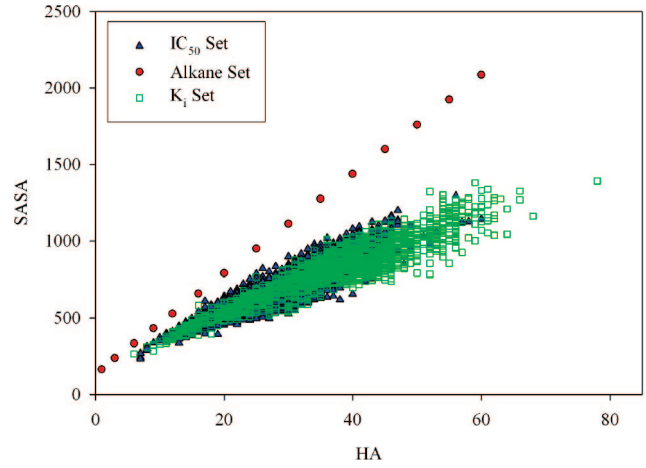

Figure 11. SASA term as a function of the heavy atom count (HA) for the model alkane set compared to the IC_{50} and K_i data sets.

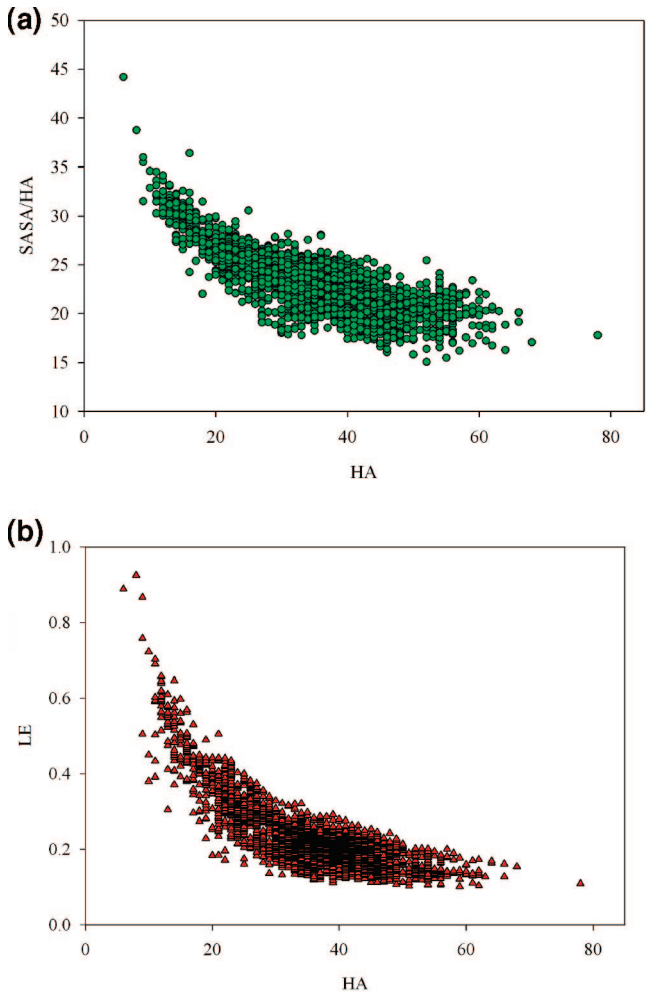

Figure 12. (a) SASA/HA term shows a close correspondnece to the (b) LE term for the K_i data set.

atoms. This means that small ligands can present more surface area for interaction per heavy atom than large ligands. Moreover, the alkane series is a best-case scenario. Complex "drug-like" molecules have more buried surface area due to branching. This is illustrated in the K_i and IC_{50} data sets where there is a more significant reduction in the per atom SASA terms as the atom count increases (Figure 11). While the very small ligands (<10 heavy atoms) fall close to the alkane SASA line, the larger ligands (>20 heavy atoms) fall well below the alkanes. Comparison of the ligand efficiencies to the SASA/heavy-atom ratios (Figure 12) show a remarkable correspondence, and provide a strong argument that one of the primary driving forces behind the systematic decline in maximal (or average) ligand efficiencies with increasing molecular size is the reduced effective surface area for the larger compounds.

Entropy. It is logical that entropy should play a role in modulating ligand efficiency.[17] But it is difficult a priori to say what or how large the entropy effect might be. One might crudely estimate the energetic cost due to the loss of conformational flexibility in the ligand upon binding by evaluating the number of accessible conformations within a certain range of the global minimum. We adopted a procedure reported by Jorgensen and Tirado-Rives[18] for estimating the loss of conformational entropy upon binding. This method entails Monte Carlo sampling followed by calculation of a conformational partition function for the bound and unbound states. In our case, we do not know the bound state and cannot estimate the difference in strain energy between bound and free ligand, but we can make the assumption that the bound state conformationally locks the ligand into a single conformation.

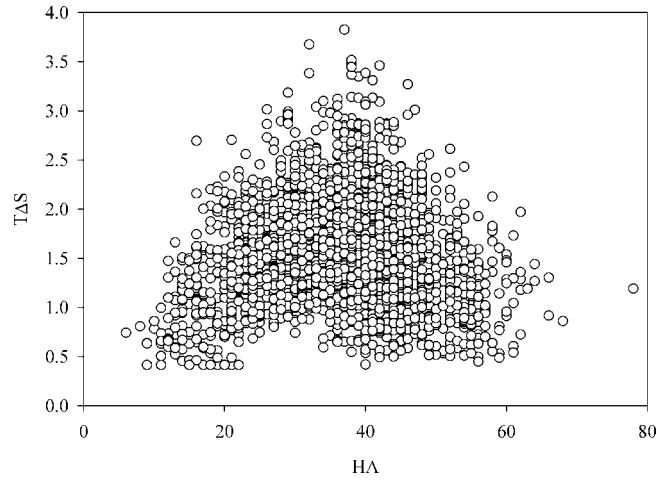

Figure 13. Entropy term plotted against the heavy atom count (HA) for the K_i set.

We calculated the conformational entropies by subjecting each ligand to 2000 steps of Monte Carlo rotational bond sampling followed by minimization. The conformational entropy was then calculated as a Boltzman weighted average of the resulting low energy conformations. The ligand conformational entropy contributions for the K_i data set are plotted against size in Figure 13. In general, the entropy values range over a wider spread as the ligands become larger. This is probably driven by the fact that at least some of the larger ligands have more rotatable bonds. This trend trails off at ~40 atoms, but our ability to reasonably sample the available conformations is greatly diminished past 10–15 rotatable bounds. Therefore, one should exercise care in drawing conclusions for the very large ligands. Figure 13 also shows that even for quite large ligands it is possible to identify structures that have small conformational entropies. For example, a large but conformationally rigid ligand should enjoy a relatively small entropy penalty.

On the basis of this analysis, the contribution of conformational entropy to the observed loss of efficiency as ligands become progressively larger is not clear-cut. On average the larger ligands incur a greater entropy cost, but there are examples of structures that are very large and have entropy values comparable to the smallest ligands in this data set. Thus, while entropy is likely a factor for some ligands it does not appear to be as directly influenced by the number of atoms as total SASA and overall ligand complexity.

Conclusion

Ligand efficiency cannot be applied across all molecular size ranges without some type of normalization for size. The maximal (and average) ligand efficiencies show a marked decline with increasing molecular size, particularly below 20 heavy atoms. This effect must be considered when comparing ligand efficiencies between ligands of significantly different size. An alternative is to adopt a size-independent metric such as the fit quality score (FQ) described above. Such a size-independent metric might be particularly important for evaluating the ligand efficiency of high-throughput hits or in a fragment-based drug design program where the ligands tend to be small. In any phase of drug discovery the fit quality score provides a simple method for directly measuring how optimally a ligand binds relative to other ligands of any size.

There are several physical effects that contribute to this observed trend. First, the available surface area for making favorable protein–ligand interactions is not linear with regard to the number of atoms. The surface area available for making favorable interactions with a protein active site in small ligands is much larger on a per atom basis than the available surface area of large ligands. This is because much of the atomic surface area in larger ligands is buried and therefore unavailable. The available surface area may well be the single most significant factor driving the reduction in ligand efficiencies as size increases. In addition, larger, more complex ligands are expected to have less optimal binding than smaller less complicated ligands. This is a consequence of the complexity of the protein–ligand interface, the greater number of constraints that must be satisfied, and the resulting structural compromises. Perhaps surprisingly, the conformational entropy penalty for ligand binding appears to play a lesser role. In any event, all of the factors we evaluated tend to contribute a headwind with regard to the ligand efficiency of larger ligands. Therefore, it is not surprising that ligand efficiencies for large ligands are always lower than for corresponding small molecules.

References

(1) Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Technol.* **2004**, *1* (4), 337–341.
(2) Abad-Zapatero, C. Ligand efficiency indices for effective drug discovery. *Expert Opin. Drug Discovery* **2007**, *2* (4), 469–488.
(3) Abad-Zapatero, C.; Metz James, T. Ligand efficiency indices as guideposts for drug discovery. *Drug Discovery Today* **2005**, *10* (7), 464–9.
(4) Andrews, P. R.; Craik, D. J.; Martin, J. L. Functional group contributions to drug-receptor interactions. *J. Med. Chem.* **1984**, *27* (12), 1648–57.
(5) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. The maximal affinity of ligands. *Proc. Natl Acad. Sci. U.S.A.* **1999**, *96* (18), 9997–10002.
(6) Reynolds, C. H.; Bembenek, S. D.; Tounge, B. A. The role of molecular size in ligand efficiency. *Bioorg. Med. Chem. Lett.* **2007**, *17* (15), 4258–4261.
(7) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* **2007**, *35*, D198–D201.
(8) Schrodinger, L. L. C. Maestro Version 8.0, 1999–2007.
(9) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110* (6), 1657–66.
(10) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A* **1997**, *101* (16), 3005–3014.
(11) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112* (16), 6127–9.
(12) Chang, G.; Guida, W. C.; Still, W. C. An internal-coordinate Monte Carlo method for searching conformational space. *J. Am. Chem. Soc.* **1989**, *111* (12), 4379–86.
(13) Jorgensen, W. L.; Duffy, E. M. Prediction of drug solubility from structure. *Adv. Drug Delivery Rev.* **2002**, *54* (3), 355–366.
(14) Systat Software, Inc. San Jose, CA 95110.
(15) $K_i = IC_{50}/(1 + ([S]/K_m))$, where [S] = concetration of substrate, and K_m is the affinity of the substrate.
(16) Hann, M. M.; Leach, A. R.; Harper, G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (3), 856–864.
(17) Murray, C. W.; Verdonk, M. L. The consequences of translational and rotational entropy lost by small molecules on binding to proteins. *J. Comput.-Aided Mol. Design* **2002**, *16* (10), 741–753.
(18) Tirado-Rives, J.; Jorgensen, W. L. Contribution of Conformer Focusing to the Uncertainty in Predicting Free Energies for Protein-Ligand Binding. *J. Med. Chem.* **2006**, *49* (20), 5880–5884.

JM701255B