6752 J. Med. Chem. 2009, 52, 6752–6756
DOI: 10.1021/jm901241e

Journal of Medicinal Chemistry Article

# Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success

Frank Lovering,*,† Jack Bikker,‡ and Christine Humblet§

*Wyeth Research, Chemical Sciences, †200 Cambridgepark Drive, Cambridge, Massachusetts 02140, ‡401 North Middletown Road, Pearl River, New York 10965, and §865 Ridge Road, Monmouth Junction, New Jersey 08543*

*Received August 19, 2009*

The medicinal chemistry community has become increasingly aware of the value of tracking calculated physical properties such as molecular weight, topological polar surface area, rotatable bonds, and hydrogen bond donors and acceptors. We hypothesized that the shift to high-throughput synthetic practices over the past decade may be another factor that may predispose molecules to fail by steering discovery efforts toward achiral, aromatic compounds. We have proposed two simple and interpretable measures of the complexity of molecules prepared as potential drug candidates. The first is carbon bond saturation as defined by fraction $sp^3$ ($Fsp^3$) where $Fsp^3$ = (number of $sp^3$ hybridized carbons/total carbon count). The second is simply whether a chiral carbon exists in the molecule. We demonstrate that both complexity (as measured by $Fsp^3$) and the presence of chiral centers correlate with success as compounds transition from discovery, through clinical testing, to drugs. In an attempt to explain these observations, we further demonstrate that saturation correlates with solubility, an experimental physical property important to success in the drug discovery setting.

## Introduction

Since Lipinski's seminal paper[1] introducing the "Rule of Five" (RO5), the medicinal chemistry community has become increasingly cognizant of the physical properties of potential drug candidates. Subsequent reports[2] have identified additional properties such as topological polar surface area (PSA) and rotatable bonds that play a role in the success of compounds transitioning from pre-exploratory to drug status. These properties have not only been incorporated in the medicinal chemistry lexicon, but are also routinely used in ADME prediction models.[3−5] As a result, properties such as molecular weight (MW), PSA, rotatable bonds, hydrogen bond donors, and hydrogen bond acceptors are scrutinized as compounds progress from hits through leads to drug candidates.

What is not directly addressed when using these descriptors is the complexity of the molecules. Over the past decade, a "movement", coined diversity oriented synthesis, has evolved with the stated aim to prepare diverse, architecturally more complex molecules.[6−10] The rationale for this is that these molecules will be more natural product-like[6] and/or more amenable to exploring additional areas of chemical space.[8] As many drugs are derived from natural products, creating more complex, drug-like libraries may offer an increased chance of finding bioactive compounds. Preparation and SAR exploration of such architecturally complex molecules will require continuing research into facile synthetic methods and more efficient ways to control regioisomerism and stereoisomerism.

Different approaches to describe the molecular complexity of molecules have been reported. Bertz[11] introduced an approach based on graph theory, whereas Barone et al.[12] adopted a more empirical approach. Each of these approaches has been applied to the analysis of synthetic intermediates leading to the total synthesis of natural products. More recently, Allu et al.[13] devised an algorithm that takes into account features such as number of rings, geminal substitutions, as well as chiral centers and saturation, while attempting to reduce the correlation of complexity to molecular weight. Schuffenhauer et al.[14] has taken a fingerprint-based approach based on atom triplets and found a relationship between complexity and biological activity.[15,16] While biological activity is correlated with greater complexity,[16] part of this is due simply to larger molecules having greater affinity for targets.[17]

Some of the same descriptors that have been applied to molecular complexity have also been used to develop natural product-like scores. Feher et al. looked at a number of descriptors, including the number of chiral centers and saturation.[18] They found that, while the number of chiral centers was an important descriptor for differentiating natural products, saturation was less so. Stahura et al. identified a small set of descriptors for a natural product-like score that included the number of single, double, and aromatic bonds. Building on this work, we considered whether we could use the fraction of saturated carbons within a molecule as a descriptor of complexity.

The rationale for looking at saturation as a key descriptor for complexity is intuitively straightforward. Saturation allows the preparation of architecturally more complex molecules resulting in the exploration of more diverse chemical space, *without increasing molecular weight significantly*. Figure 1 shows the various isomers of dimethylpyridine versus dimethylpiperidine. While saturation results in a slight increase in MW due to the addition of six protons, it allows access to significantly more isomers. Five dimethylpyridine isomers are accessible versus 34 isomers of dimethylpiperidine.

*To whom correspondence should be addressed. Frank Lovering, e-mail flovering@wyeth.com, phone (617) 665-5612.

 Published on Web 10/14/2009 

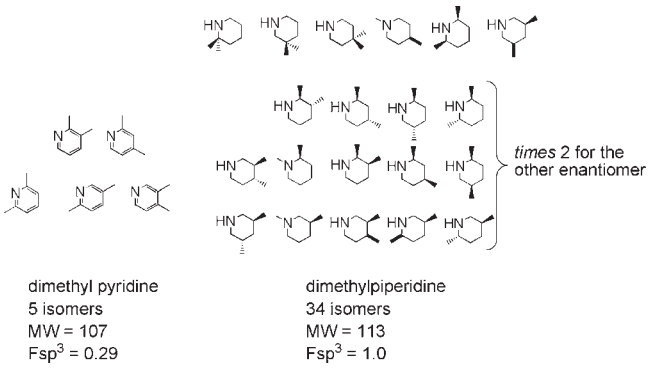


**Figure 1.** Isomers of dimethylpyridine and dimethylpiperidine.

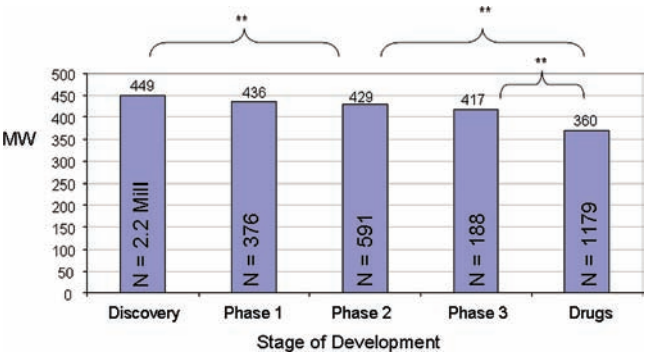


**Figure 2.** Mean molecular weight for compounds in different stages of development. $^{**}P$ value $< 0.001$.

Not only is there access to more isomers via saturation, the compounds have greater three-dimensionality than the pyridine counterparts.

In addition to the potential for more diversity for a given molecular weight, we also hypothesized that increasing $sp^3$ character may improve several molecular attributes that contribute to clinical success. The increased opportunity to design in out-of-plane substituents and to adjust molecular shape could increase receptor/ligand complimentarity. This might allow the engineering of additional protein−ligand interactions not accessible to a flat aromatic ring, and thus improve potency and selectivity to a given target which should mitigate off-target effects. While aromatic features can provide an opportunity to develop $\pi-\pi$ interactions[19] or $\pi$−cation interactions,[20] an overall level of saturation may provide the molecule with an opportunity to better place these types of moieties. Furthermore, there is considerable medicinal chemistry lore[21] that suggests that reducing the aromatic character of a molecule might improve physical characteristics, such as solubility. Two recent studies have identified the aromatic proportion of a molecule as a key descriptor to predict solubility.[22,23] Notably, any compound with an intravenous mode of delivery should necessarily be more soluble. If increased $sp^3$ character led to better clinical success, we hypothesized that this would be evident if we compared marketed drugs, compounds that successfully passed stages of clinical testing, and all drug-like molecules synthesized for drug targets.

Like complexity and natural product-likeness, saturation can be calculated using various descriptors including the number of aryl, double, and single bonds or the number of aromatic rings. Badertscher et al.[24] developed a more elaborate formalism to calculate saturation. We felt that the best approach was a simple measure of saturation that is readily interpretable (eq 1).

$$\mathrm{Fsp}^3 = (\text{number of } sp^3 \text{ hybridized carbons}/\text{total carbon count}) \quad (1)$$

We set out to determine whether there was historical evidence that increased saturation improved the likelihood of a compound becoming a drug. The approach taken here is akin to that of Wenlock et al.[25] in their analysis of property profiles for development and marketed drugs. The GVK BIO database[26] was used to source the stage of development of compounds (discovery, phase 1, 2, and 3, and drugs). Three types of descriptors were calculated. The first was a measure of the degree of saturation of each compound (Fsp3 from eq 1 above). The second attempted to capture the presence of stereo centers, and the third was molecular weight. For each category of clinical progression, the average saturation, chiral center counts, and molecular weights were calculated. We demonstrate that there is a significant correlation between both increasing saturation and increasing presence of chiral centers as compounds progress through clinical testing, suggesting that this process increases the enrichment of each. Furthermore, we demonstrate that the proposed descriptor, $Fsp^3$, does indeed correlate with two physical properties: solubility and melting point.

## Methods

**Discovery and Clinical Data.** All of the compounds studied were retrieved from the GVK BIO database. The phase reported for a compound is the highest phase a compound reached (1, 2, or 3). The purpose of this study is to determine if compounds prepared to become drugs have a greater chance of success if the compound is more complex as reflected by its saturation. Very few compounds are reported as phase 1, 2, or 3 prior to 1980.[27] Therefore, all compounds where GVK BIO had a reported publication date prior to 1980 were removed from the study (phase 1, 2 compounds; phase 2, 6 compounds; phase 3, 1 compounds; drugs, 1078 compounds). Discovery compounds are all compounds reported (1980 and onward) as either having biological activity or reported in a biologically relevant patent. Compounds were required to have at least four carbons and molecular weight under 1000 Da.

**Calculation of Properties.** Pipeline Pilot 7.5[28] was used to calculate the following properties: $sp^3$ hybridized carbons, molecular weight, and rotatable bonds. The number of stereo centers was also calculated using Pipeline Pilot 7.5 by summing Num_UnknownTrueStereoAtoms and the Num_TrueStereoAtoms.

Student's $t$ test was applied to $Fsp^3$ and molecular weight as a function of the stage of development to determine whether the differences of any two means were statistically significant using JMP 8.0.[29]

**Solubility Data.** Data from Hou et al.[30] was retrieved from a public Website (http://modem.ucsd.edu/adme/databases/databases_logS.htm). Of the 1290 compounds, 1202 had 4 or more carbons and molecular weight under 1000 Da. The compounds were binned such that each bin number contained compounds with $\log S \pm 1$.

**Melting Point Data.** Data from Karthikeyan et al.[31] was downloaded from http://cheminformatics.org/. The data contained 4450 compounds of which 4445 were converted to structures by pipeline Pilot 7.5. Of the 4445 compounds, 4432 had 4 or more carbons and molecular weight under 1000 Da. The melting point data were binned such that each bin contained compounds with melting point ±25.

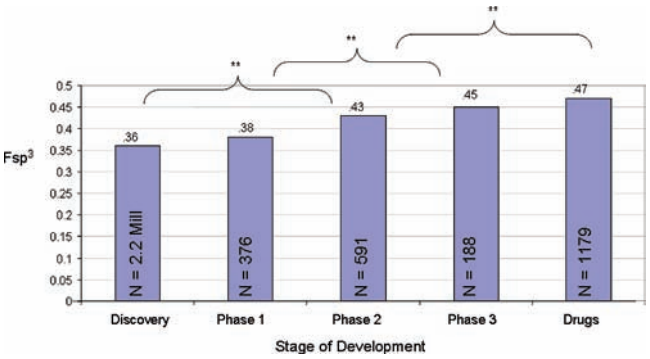


**Figure 3.** Mean $Fsp^3$ for compounds in different stages of development. $^{**}P$ value $< 0.001$.

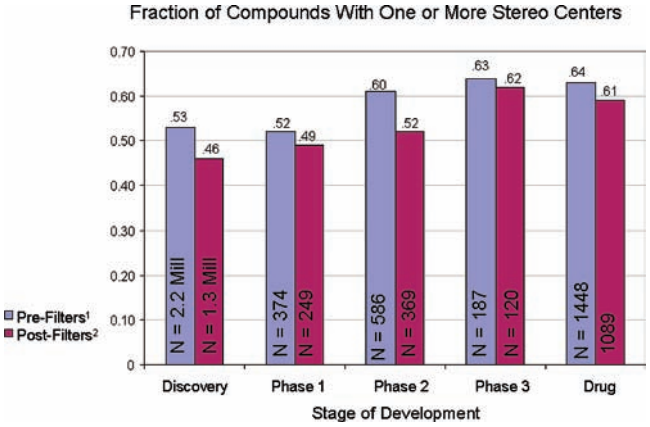


**Figure 4.** Fraction of compounds with one or more stereo centers. [1]Fraction of compounds that have one or more stereo centers. [2]Fraction of compounds that have one or more stereo centers after removal of all compounds that failed any of the RO5 or have greater than 10 rotatable bonds.

## Results and Discussion

To establish that our drug and clinical candidates database contained the same trends as those previously reported, the trend of MW through clinical progression was investigated (Figure 2). As Wenlock et al. reported,[25] molecular weight went down at each stage from discovery to drug (21.4% from discovery to drugs). The trend was statistically significant between any stages separated by a stage except from phase 1 to phase 3. Interestingly, the average drug molecular weights are somewhat higher than those previously reported.[1,25] This is likely due to two factors. The data set utilized herein is focused on compounds that have at least four carbons, eliminating very small molecules. Also, compounds with publications reported by GVK BIO[26] prior to 1980 have been removed. Thus, the present set of compounds better reflects more recent trends of the drug discovery process.

Using this database, we then demonstrated that a trend emerged when we applied our complexity measurement to clinical progression. The average $Fsp^3$ was 0.36 for discovery compounds and increased to 0.47 for drugs (Figure 3). This represents a 31% increase in the saturated fraction. Importantly, the trend is carried through all of the stages from discovery to drug where each phase had a higher $Fsp^3$. The change in average $Fsp^3$ was statistically significant between adjacent stages in only one case (phase 1 to phase 2). However, the change is statistically significant between any stages separated by a stage (discovery to phase 2, phase 1 to phase 3, phase 2 to drug) illustrating how compounds with greater saturation are more likely to succeed at each stage from discovery to drug.

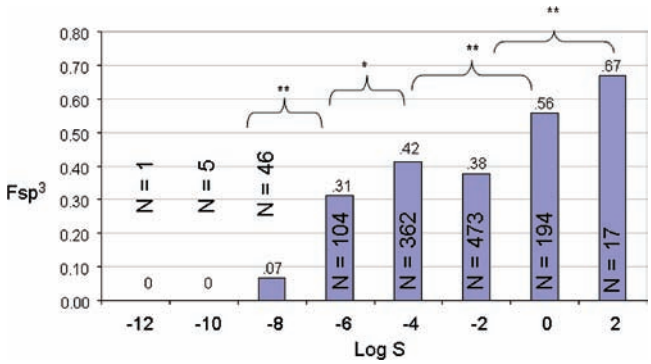


**Figure 5.** $Fsp^3$ as a function of log *S*. $^{*}P$ value $< 0.01$. $^{**}P$ value $< 0.001$.

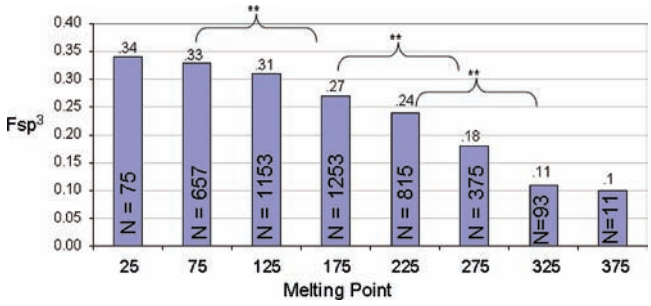


**Figure 6.** $Fsp^3$ as a function of melting point. $^{**}P$ value $< 0.001$.

A second descriptor that we proposed to capture molecular complexity was the presence of stereo centers. Figure 4 shows the percentage of compounds that had one or more stereo centers at any stage of development. 53% of the discovery compounds had one or more stereo centers. While phase 1 had about the same percentage, 60% of phase 2 compounds and 64% of drugs had one or more stereo centers, a 21% increase over discovery. To confirm that compounds with undesirable properties were not overly influencing the outcome, all compounds that violated any one of the RO5, as well as those that had $> 10$ rotatable bonds were removed from the analysis. In this case, 46% of discovery compounds had one or more stereo centers. As we transition from discovery to drug, we see an increase in the percentage of compounds that have one or more stereo centers. 61% of drugs had compounds with one or more stereo centers, a 33% increase.

The central premise on which we have relied was that greater saturation would allow greater complexity and thus access to more of the available chemical space. However, saturation will also affect physical properties. Toward this end, we investigated whether our measure of saturation was correlated with solubility and melting point.

Our complexity metric is correlated with both solubility and melting point when applied to literature data sets. $Fsp^3$ was calculated on 1202 compounds derived from a solubility data set previously reported by Hou et al.[30] and 4432 compounds from a melting point data set from Karthikeyan et al.[31] As can be seen in Figure 5, the average $Fsp^3$ went up with log *S*. A relationship between melting point and $Fsp^3$ is also seen in Figure 6 where the average $Fsp^3$ is found to decrease with increasing melting point. Given the relationship between solubility and $Fsp^3$, this finding is not surprising. Yalkowsky

and Valvani[32] and later Jain and Yalkowsky[33] reported the general solubility equation where solubility can be estimated based on melting point and log *P*. The influence of saturation on melting point has been realized for years due to the impact of hydrogenation on the melting point of oils.[34] Moreover, melting points have been utilized to better predict drug absorption.[35] Thomas et al.[36] suggested that highly ordered crystal lattices resulting in compounds with a melting point above 250 °C will negatively impact oral bioavailability.

## Conclusion

More highly complex molecules, as measured by saturation, have the capacity to access greater chemical space. This results in greater potential to identify compounds that better complement the spatial subtleties of target proteins. Importantly, the three-dimensionality that saturation imparts may also result in greater selectivity, resulting in fewer off-target effects. Herein, we have identified a very simple descriptor for saturation which is easily interpretable. As compounds are prepared in the drug discovery setting and transition from discovery through clinical trials to drugs, those that are more highly saturated are more likely to succeed in these transitions. Saturation also increased the likelihood of higher solubility and lower melting points. Compounds are much more likely to succeed as drugs if they have appropriate values for these properties. Another descriptor for complexity, the presence of stereo centers, also increases as we transition from discovery, through clinical trials, to drugs. This held true even after filtering compounds that failed any one of the RO5, as well as had greater than 10 rotatable bonds.

Presently, the emphasis in the pharmaceutical industry stresses the efficient assembly of molecules, often in a parallel manner. Advances over the last 10−15 years in the coupling of $sp^2-sp^2$ carbons,[37,38] as well as other $sp^2$ couplings,[39,40] have made the preparation of molecules with greater unsaturation particularly amenable to parallel synthesis. While these advances have contributed to drug discovery, they have also biased efforts at the bench. Diversity oriented synthesis has sought to reverse this trend by identifying facile syntheses of more complex molecules. The results presented here give impetus to this movement.

**Acknowledgment.** The authors thank Rajiah Denny for kind help with the GVK BIO database and Ray Unwalla for helpful discussions.

**Editor's Note.** During the revision process, a paper was published that demonstrates that larger numbers of aromatic rings negatively affect several drug-like properties.[41]

## References

(1) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev*. **1997**, *23*, 3–25.
(2) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem*. **2002**, *45*, 2615–2623.
(3) Gleeson, M. P. Generation of a set of simple, interpretable ADMET rules of thumb. *J. Med. Chem*. **2008**, *51*, 817–834.
(4) Martin, Y. C. A bioavailability score. *J. Med. Chem*. **2005**, *48*, 3164–3170.
(5) Pajouhesh, H.; Lenz, G. R. Medicinal chemical properties of successful central nervous system drugs. *NeuroRx* **2005**, *2*, 541–553.
(6) Arya, P.; Joseph, R.; Gan, Z.; Rakic, B. Exploring new chemical space by stereocontrolled diversity-oriented synthesis. *Chem. Biol*. **2005**, *12*, 163–180.
(7) Morton, D.; Leach, S.; Cordier, C.; Warriner, S.; Nelson, A. Synthesis of natural-product-like molecules with over eighty distinct scaffolds. *Angew. Chem., Int. Ed*. **2009**, *48*, 104–109.
(8) Burke, M. D.; Schreiber, S. L. A planning strategy for diversity-oriented synthesis. *Angew. Chem., Int. Ed*. **2004**, *43*, 46–58.
(9) Schreiber, S. L. Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* **2000**, *287*, 1964–1969.
(10) Tan, D., S. Diversity-oriented synthesis: exploring the intersections between chemistry and biology. *Nat. Chem. Biol*. **2005**, *1*, 74–84.
(11) Bertz, S. H. The first general index of molecular complexity. *J. Am. Chem. Soc*. **1981**, *103*, 3599–3601.
(12) Barone, R.; Chanon, M. A New and simple approach to chemical complexity. application to the synthesis of natural products. *J. Chem. Inf. Comp. Sci*. **2001**, *41*, 269–272.
(13) Allu, T. K.; Oprea, T. I. Rapid evaluation of synthetic and molecular complexity for in silico chemistry. *J. Chem. Inf. Mod*. **2005**, *45*, 1237–1243.
(14) Schuffenhauer, A.; Floersheim, P.; Acklin, P.; Jacoby, E. Similarity metrics for ligands reflecting the similarity of the target proteins. *J. Chem. Inf. Comp. Sci*. **2003**, *43*, 391–405.
(15) Schuffenhauer, A.; Brown, N.; Selzer, P.; Ertl, P.; Jacoby, E. Relationships between molecular complexity, biological activity, and structural diversity. *J. Chem. Inf. Mod*. **2006**, *46*, 525–535.
(16) Selzer, P.; Roth, H.-J.; Ertl, P.; Schuffenhauer, A. Complex molecules: do they add value? *Curr. Opin. Chem. Biol*. **2005**, *9*, 310–316.
(17) Hajduk, P. J. Fragment-based drug design: how big is too big? *J. Med. Chem*. **2006**, *49*, 6972–6976.
(18) Feher, M.; Schmidt, J. M. Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem. Inf. Comp. Sci*. **2003**, *43*, 218–227.
(19) McGaughey, G. B.; Gagne, M.; Rappe, A. K. Pi-stacking interactions: alive and well in proteins. *J. Biol. Chem*. **1998**, *273*, 15458–15463.
(20) Dougherty, D. A. Cation-pi interactions in chemistry and biology: a new view of benzene, phe, tyr, and trp. *Science* **1996**, *271*, 163–168.
(21) Ando, H. Y.; Radebaugh, G. W., Property-based drug design and preformulation. In *Remington: The Science and Practice of Pharmacy*; Lippincott, Williams and Witkins; Philadelphia, 2005; pp 720−744.
(22) Delaney, J. S. ESOL: estimating aqueous solubility directly from molecular structure. *J. Chem. Inf. Comp. Sci*. **2004**, *44*, 1000–1005.
(23) Lamanna, C.; Bellini, M.; Padova, A.; Westerberg, G.; Maccari, L. Straightforward recursive partitioning model for discarding insoluble compounds in the drug discovery process. *J. Med. Chem*. **2008**, *51*, 2891–2897.
(24) Badertscher, M.; Bischofberger, K.; Munk, M. E.; Pretsch, E. A. Novel formalism to characterize the degree of unsaturation of organic molecules. *J. Chem. Inf. Comp. Sci*. **2001**, *41*, 889–893.
(25) Wenlock, M. C.; Austin, R. P.; Barton, P.; Davis, A. M.; Leeson, P. D. A comparison of physiochemical property profiles of development and marketed oral drugs. *J. Med. Chem*. **2003**, *46*, 1250–1256.
(26) GVK Bio web site [online], http://www.gvkbio.com/Informatics.html/.
(27) Some compounds did not have a publication date recorded and these were kept in the analysis: 28 phase 1, 50 phase 2 and 10 phase 3.
(28) Pipeline Pilot is available from Accelrys, http://accelrys.com/products/scitegic/.
(29) JMP is available from http://www.jmp.com/.
(30) Hou, T. J.; Xia, K.; Zhang, W.; Xu, X. J. ADME evaluation in drug discovery. 4. Prediction of aqueous solubility based on atom contribution approach. *J. Chem. Inf. Comp. Sci*. **2003**, *44*, 266–275.
(31) Karthikeyan, M.; Glen, R. C.; Bender, A. General melting point prediction based on a diverse compound data set and artificial neural networks. *J. Chem. Inf. Mod*. **2005**, *45*, 581–590.
(32) Yalkowsky, S. H.; Valvani, S. C. Solubility and partitioning I: Solubility of nonelectrolytes in water. *J. Pharm. Sci*. **1980**, *69*, 912–922.
(33) Neera, J.; Samuel, H. Y. Estimation of the aqueous solubility I: Application to organic nonelectrolytes. *J. Pharm. Sci*. **2001**, *90*, 234–252.
(34) *Handbook of Chemistry and Physics*, 72nd ed., CRC Press: Boca Raton, FL, 1991.
(35) Chu, K. A.; Yalkowsky, S. H. An interesting relationship between drug absorption and melting point. *Int. J. Pharm*. **2009**, *373*, 24–40.

(36) Thomas, V. H.; Bhattachar, S.; Hitchingham, L.; Zocharski, P.; Naath, M.; Surendran, N.; Stoner, C. L.; El-Kattan, A. The road map to oral bioavailability: an industrial perspective. *Exp. Opin. Drug Metab. Tox.* **2006**, *2*, 591–608.
(37) Adam, F. L.; Gregory, C. F. Palladium-catalyzed coupling reactions of aryl chlorides. *Angew. Chem., Int. Ed.* **2002**, *41*, 4176–4211.
(38) Shawn, D. W.; Timothy, E. B.; Joseph, R. M.; Stephen, L. B. A rationally designed universal catalyst for suzuki-miyaura coupling processes13. *Angew. Chem., Int. Ed.* **2004**, *43*, 1871–1876.
(39) Huang, X.; Anderson, K. W.; Zim, D.; Jiang, L.; Klapars, A.; Buchwald, S. L. Expanding Pd-Catalyzed Bond-Forming Processes: The First Amidation of Aryl Sulfonates, Aqueous Amination, and Complementarity with Cu-Catalyzed Reactions. *J. Am. Chem. Soc.* **2003**, *125*, 6653–6655.
(40) Yin, J.; Zhao, M. M.; Huffman, M. A.; McNamara, J. M. Pd-catalyzed N-arylation of heteroarylamines. *Org. Lett.* **2002**, *4*, 3481–3484.
(41) Ritchie, T. J.; Macdonald, S. J. F. The impact of aromatic ring count on compound developability - are too many aromatic rings a liability in drug design. *Drug Disc. Today* **2009**, doi:10.1016/j.drudis.2009.07.014.