Investigation of the Relationship between Topology and Selectivity for Druglike Molecules

Yidong Yang,*,[†] Hongming Chen,[†] Ingemar Nilsson,[‡] Sorel Muresan,[†] and Ola Engkvist*,[†]

[†]*DECS Global Compound Sciences and* [‡]*Department of Medicinal Chemistry, AstraZeneca R&D Mölndal, SE-431 83 Mölndal, Sweden*

There is a strong interest in drug discovery and development to advance the understanding of pharmacological promiscuity. Improved understanding of how a molecular structure is related to promiscuity could help to reduce the attrition of compounds in the drug discovery process. For this purpose, a descriptor is introduced that describes the structural complexity of a compound based on the size of its molecular framework (MF) in relation to its overall size. It is defined as the fraction of the size of the molecular framework versus the size of the whole molecule (f_{MF}). It is demonstrated that promiscuity correlates with f_{MF} for large f_{MF} values. The observed correlation is not due to lipophilicity. To provide further explanation of this observation, it was found that the number of terminal ring systems in a compound is correlated with promiscuity. The analysis presented here might help medicinal chemists to improve the selectivity for compounds in drug discovery projects.

Introduction

The exponential advance made over the past decade in human genomic and proteomic research has had a big influence on the search for new therapeutic agents. The prevailing assumption is that single agents that modulate distinct therapeutically relevant targets confer better therapeutic efficacy and fewer side effects in most disease areas. Consequently, most of the drug discovery projects are focused on designing highly selective compounds. Pharmacological promiscuity is especially undesirable for this "one drug, one target" paradigm.[1−3] However, there are cases (e.g., complex diseases such as Alzheimer's disease[4] and central nervous system disorders[5]) for which the design of new medicines that address multiple targets simultaneously may be desirable. There is also evidence for promiscuity as a key contributor to the clinical efficacy of many newer anticancer drugs.[6] An understanding of the molecular and structural basis for compound promiscuity could help to design, optimize, and prioritize suitable lead structures at the earliest possible stage of a drug design project and therefore significantly increase research productivity.

In pharmaceutical research, off-target activities (secondary pharmacology) are usually inferred from in vitro testing of the compound against a panel of proteins.[7] Published computational models investigating the relationship between promiscuity and adverse effects have been based on such types of data sets.[8−10] Several recent studies showed that lipophilicity (ClogP) is a principal determinant of pharmacological promiscuity. A compound with high ClogP usually exhibits higher promiscuity than a compound with low ClogP. Other molecular properties such as ionization state, basicity, and presence of certain functional groups have also been found to influence promiscuity.[11−15] Several studies have been published on the relationship between molecular size and promiscuity. Analysis of a high-throughput screening (HTS) data set[11] showed that promiscuity decreases with higher molecular weight (MW). However, this result is in contrast with the analysis of safety pharmacology profiling data.[7] This analysis showed that molecules with higher MW are more promiscuous. An analysis by Leeson et al.[13] on the BioPrint data set and in-house cross-screening data showed no clear relationship between MW and promiscuity. Similar results were also found in another study.[15] The different conclusions indicate that a potential correlation between size and promiscuity is highly context dependent.

However, even though promiscuity may not directly link to the size of compound, it could still be influenced by the molecular topology. In this article we focus on investigating the correlation between promiscuity and a generic structural description of a compound. Applying more interpretable structural features for describing a molecule might provide further insights to its pharmacological response. For this purpose, we first divided the molecule into a molecular framework (MF[a]) and side chains as defined by Bemis and Murcko.[16,17] The descriptor f_{MF} is defined as the number of heavy atoms (Nheavy) in the MF divided by the total number of heavy atoms in the molecule as shown in eq 1, where $f_{MF} \in [0, 1]$.

$$f_{MF} = \frac{Nheavy_{MF}}{Nheavy_{total}} \quad (1)$$

A simple interpretation of this descriptor is that molecules with low f_{MF} values have relatively small frameworks and are decorated with many and/or larger side chains. On the other hand, a molecule with a high f_{MF} has a large MF and only a few side chain atoms. Compounds with $f_{MF} = 1$ have no side chains at all. Acyclic molecules ($f_{MF} = 0$) that do not contain

*To whom correspondence should be addressed. For Y.Y.: phone, +46 (0)31 70 61097; fax, +46 (0)31 77 63792; e-mail, yidong@gmail.com. For O.E.: phone, +46 (0)31 7065978; fax, +46 (0)31 77 63792; e-mail, ola.engkvist@astrazeneca.com.

[a] Abbreviations: MF, molecular framework; Nheavy, number of heavy atoms; f_{MF}, fraction of molecular framework; TR, terminal ring.
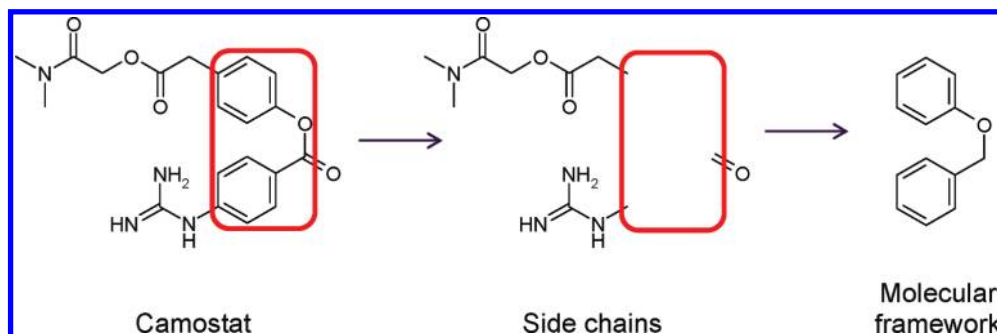
Figure 1. Disconnecting side chains from the original molecule leads to its molecular framework.
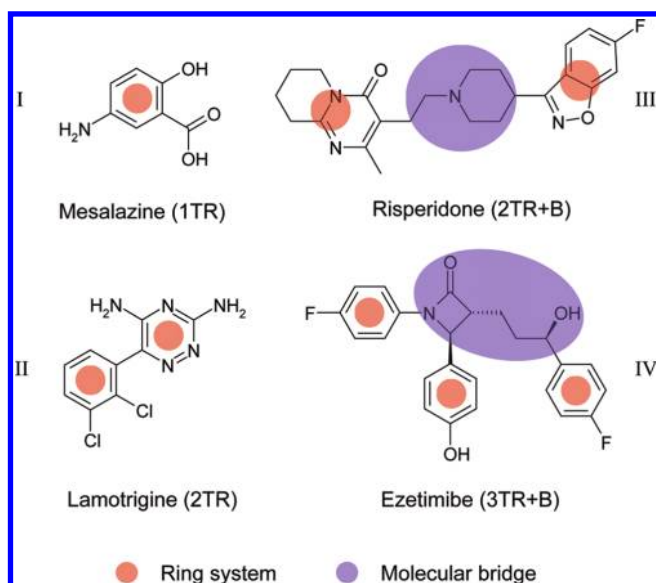

Figure 2. Examples of the four different topology classes (TR = terminal ring, B = molecular bridge).

any ring system are not included in this analysis. f_{MF} is not correlated to the number of heavy atoms ($r^2 = 0.007$ for the BioPrint data set used in this study), and the descriptor therefore describes the molecular topology in a size independent way. It is demonstrated that promiscuity increases for large f_{MF} and that the observed increase is not related to lipophilicity. To further explain this observation, the molecules were divided into four different classes based on their topology. The relationship between the four topology classes and promiscuity was analyzed.

Methods

Molecular Framework Analysis. The main descriptor used in the analysis is based on the hierarchical molecular classification scheme proposed by Bemis and Murcko.[16,17] As can be seen in Figure 1, Camostat can be fragmented into three side chains and a MF. The MF consists of two rings connected by a two-atom linker. The f_{MF} of Camostat can be calculated as the number of heavy atoms in the MF (14) divided by the total number of heavy atoms in the molecule (29); accordingly, f_{MF} is 0.483.

Since the MF is obtained by pruning all terminal side chain atoms, it consists of all the ring systems and linkers in a molecule. In an attempt to further investigate how a MF relates to promiscuity, we classified the MFs according to their ring system topology. In order to classify compounds into only a few distinct topology classes for our analysis, a more abstract topological classification scheme is used compared to other methods such as the molecular equivalence indices (MEQI).[18,19] Once the MF is
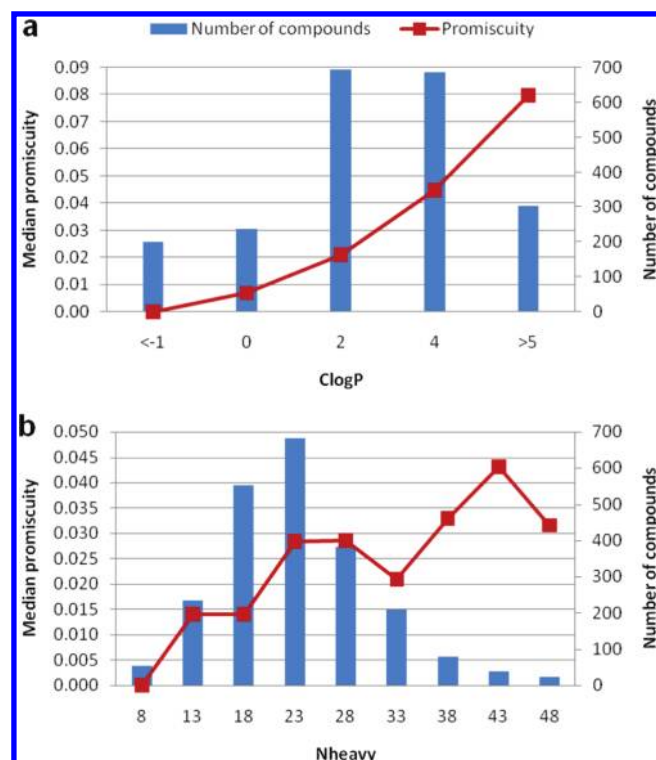

Figure 3. Relationship between median promiscuity and lipophilicity, ClogP (a), and the relationship between median promiscuity and number of heavy atoms, Nheavy (b), for 2267 compounds from the BioPrint data set.
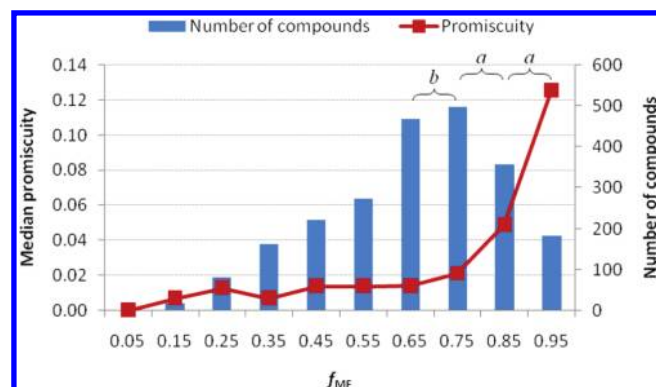

Figure 4. Relationship between the median promiscuity and f_{MF}: (a) $p < 0.0001$; (b) $p < 0.01$.

generated, ring systems connected to only one other ring system are identified and labeled as terminal rings (TRs). All other atoms between the TRs are grouped together and labeled as the
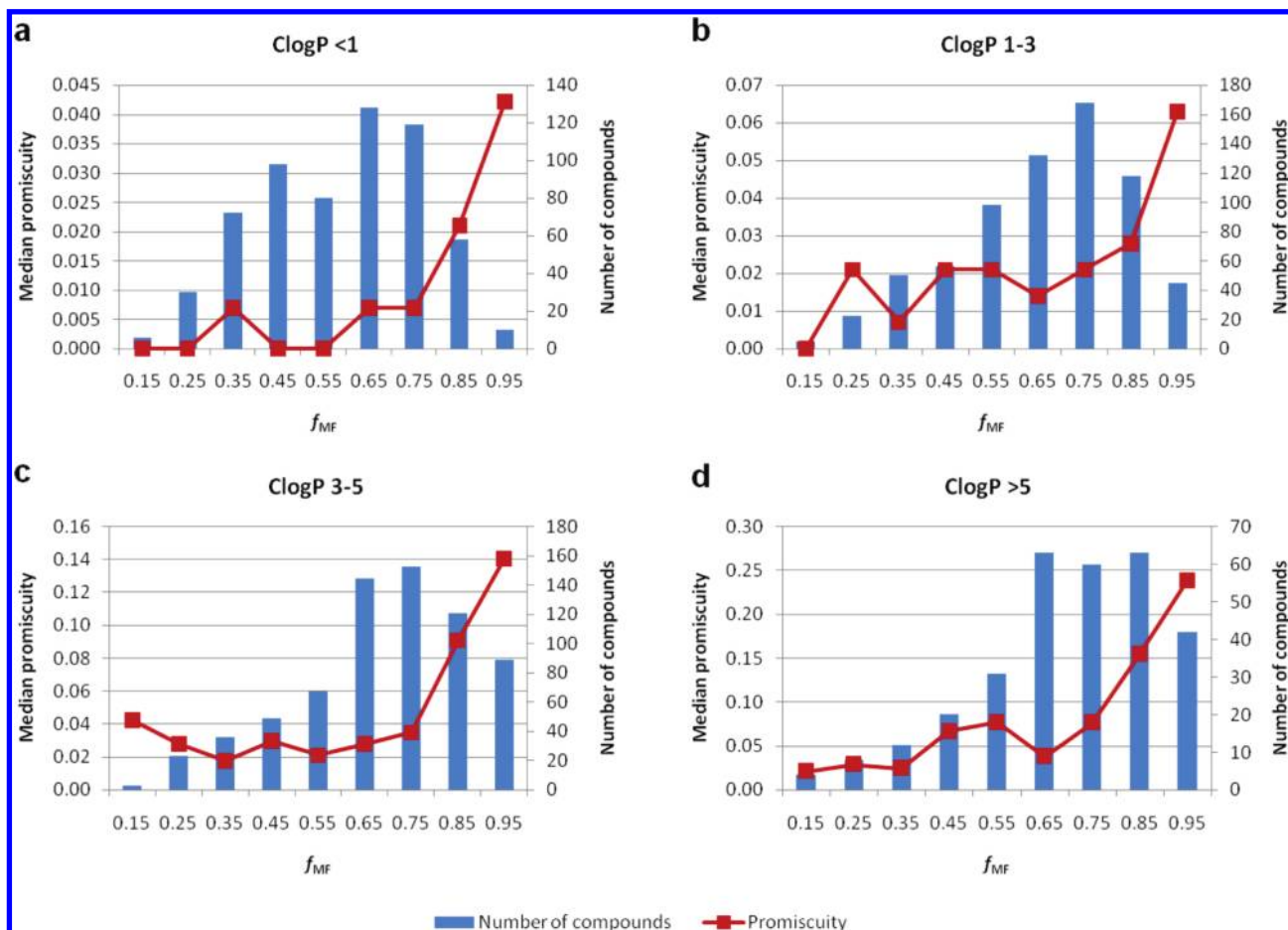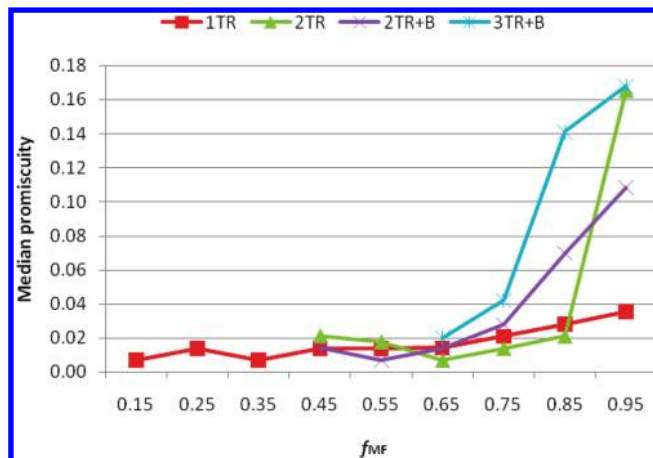
Figure 5. Median promiscuity versus f_{MF} for the four ClogP intervals. The differences between the 0.75 and 0.95 bins are statistically significant with $p < 0.0001$ for parts b, c, and d and with $p < 0.001$ for part a.
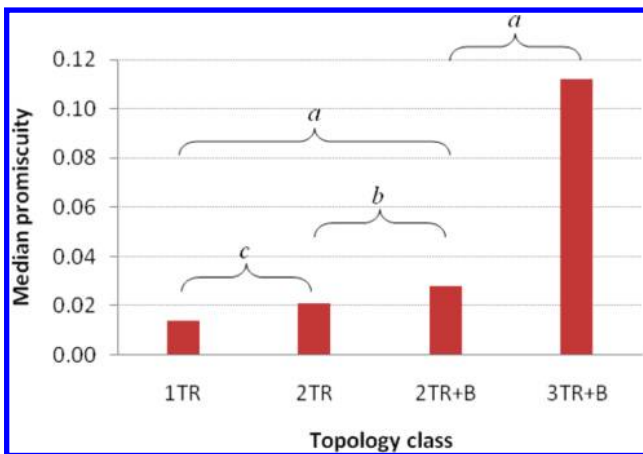
Figure 6. Median promiscuity for the four different topology classes: (a) $p < 0.0001$; (b) $p < 0.01$; (c) $p < 0.04$.

Figure 7. Median promiscuity versus f_{MF} for the four different topology classes. For the 2TR and 2TR $+$ B classes the changes are statistically significant with $p < 0.0001$ and for the 3TR $+$ B class with $p < 0.0005$ between the 0.75 and 0.95 bins. The changes seen for 1TR are not statistically significant because of too few compounds for the 0.95 bin. The number of compounds for each topology class and f_{MF} are given in Table S1.

molecular bridge. We can now classify molecules into different topological classes according to the number of terminal ring systems and the presence of a molecular bridge as shown in Figure 2. For example, compound **I** belongs to the "one terminal ring system" (1TR) class. This class contains molecules with only one ring system and no molecular bridge. Compound **II** belongs to the "two terminal ring systems" (2TR) class. This class contains molecules with two ring systems directly connected to each other and no molecular bridge. Compound **III** belongs to the 2TR $+$ B class. This class contains molecules with two terminal ring systems and a molecular bridge. Compound **IV** belongs to the 3TR $+$ B class. This class

contains molecules with three terminal ring systems and a molecular bridge. Since almost all molecules belong to these four topology classes according to our study, compounds containing four or even more terminal ring systems were not considered in this analysis.

The MF and side chains were generated using Pipeline Pilot.[20] f_{MF}, number of heavy atoms, and MW were calculated with
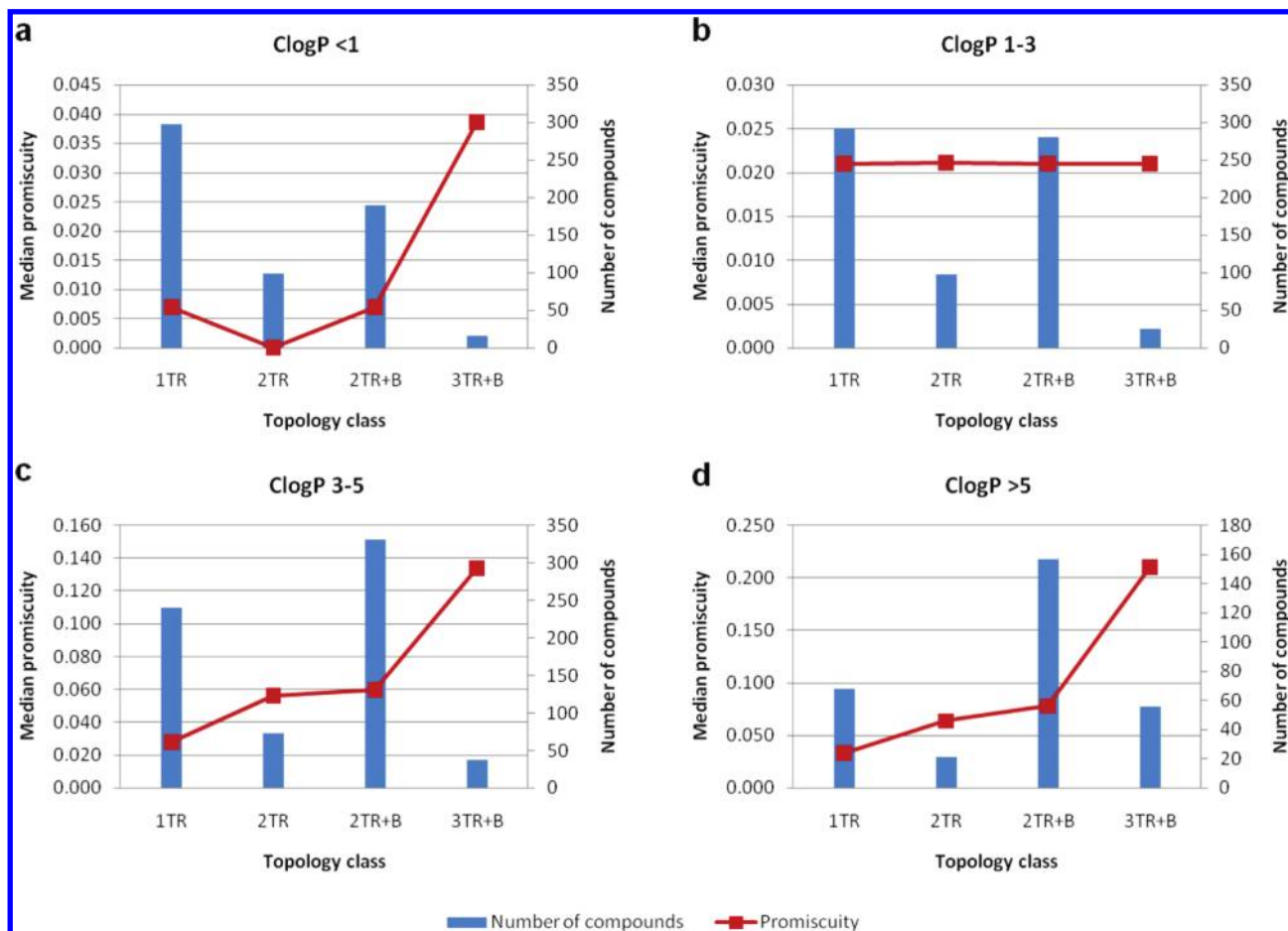
Figure 8. Median promiscuity for the four different topology classes and ClogP ranges.

Pipeline Pilot. ClogP was calculated with BioByte's ClogP program.[21] The classification of the MFs into topological classes was done with an in-house C++ program based on the OEChem Toolkit.[22]

Bioactivity Data. The analysis is based on selectivity data of 2267 compounds from the BioPrint database.[13,23] The data set consists mainly of marketed drugs, withdrawn drugs, and reference compounds. They have been tested in a panel of more than 200 diverse protein targets representing various target families like GPCRs, ion channels, transporters, and enzymes. The promiscuity score for each compound was defined as the number of targets for which the compound displayed $\geq 50\%$ inhibition (at 10 μM) divided by the total number of panel targets the compound has been screened against. Very small and large compounds (outside the range of 5−60 heavy atoms) were deemed irrelevant for this analysis and were therefore removed. The nonparametric Wilcoxon rank-sum test[24] was applied to determine whether the differences in median promiscuity were statistically significant. All statistical analyses were performed with JMP.[25]

Results and Discussion

First, the relationships between promiscuity and lipophilicity (ClogP) and between promiscuity and size (number of heavy atoms, Nheavy) were investigated for the data set. The ClogP and Nheavy values were binned and then plotted versus the median promiscuity. The results are shown in Figure 3, and they are in agreement with results from previous studies.[7,13,15] As expected, the promiscuity increases with increasing ClogP. The relationship between promiscuity and molecular size, as measured by Nheavy, shows an overall trend that larger compounds are more promiscuous. However, the trend is less pronounced

than for ClogP. Thus, the conclusion could be affected by exactly which compounds have been tested. We believe that our result is in line with the divergent conclusions for the relationship between size and promiscuity reported by other groups.[7,11,13,15]

Thereafter, how f_{MF} relates to promiscuity was investigated. The results are shown in Figure 4. Compounds with f_{MF} larger than 0.65 are significantly more promiscuous than compounds with a smaller f_{MF}. The change in median promiscuity between the f_{MF} bins is statistically significant with a $p < 0.0001$, when f_{MF} is above 0.75. To distinguish the effect of f_{MF} on promiscuity from the effect of lipophilicity, the data set was divided into four bins corresponding to ClogP of < 1, $1−3$, $3−5$, and > 5. Figure 5 shows that the same trend is observed for all the different ClogP ranges. This result confirms that f_{MF} has a unique effect on the promiscuity that is not dependent on compound lipophilicity. However, the median promiscuity is highest for compounds with ClogP > 5 and lowest for ClogP < 1, as already shown in Figure 3a. Note that because the f_{MF} represents the size of the MF in relation to the compound size, the descriptor is uncorrelated to the overall size of the compound.

In order to be able to interpret the observed promiscuity trend, the compounds were classified according to their topology as discussed in the Methods. As can be seen in Figure 6, compounds belonging to the 1TR class show the highest selectivity while compounds belonging to the 3TR + B class are the most promiscuous. The Wilcoxon rank-sum test shows that the differences in promiscuity between the topological classes are statistically significant. In Figure 7 it is shown that for all four topology classes the promiscuity
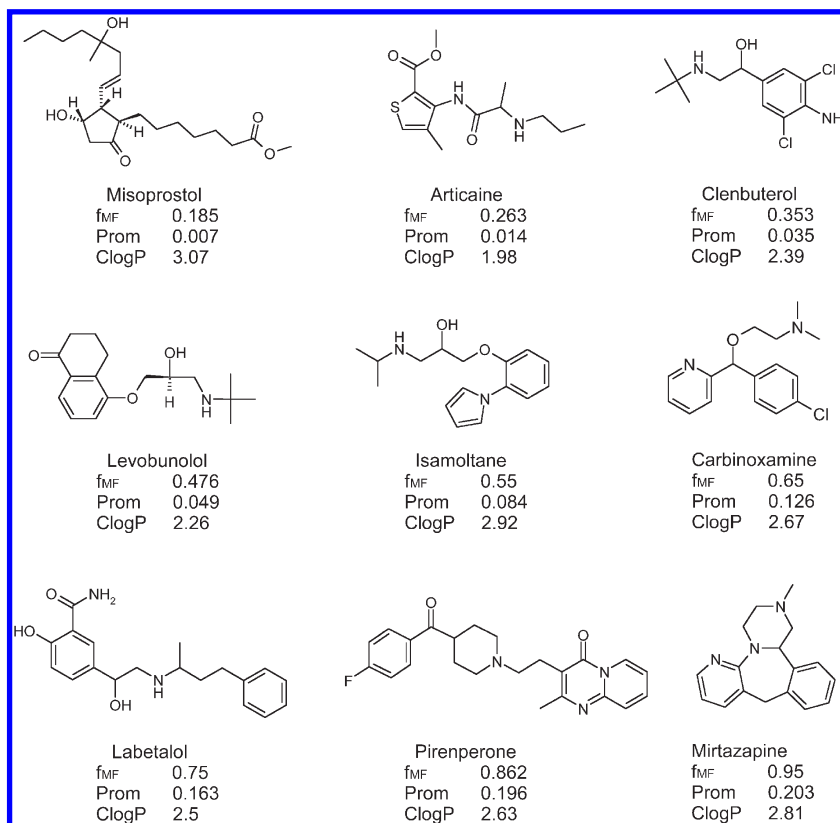
Figure 9. Examples of compounds with increasing f_{MF} (from top to bottom) and promiscuity (Prom).

increases when f_{MF} is above 0.65. However, the magnitude is very different for the different topological classes. The increase in promiscuity is small for the 1TR class and not statistically significant, while the increase is much larger for the other topology classes. These results indicate that compounds with only one ring system and many side chain atoms are on average more selective than compounds of the opposite type. Most of the compounds with a f_{MF} value larger than 0.65 belong to the 2TR + B and 3TR + B classes (Table S1 in the Supporting Information). Median values for f_{MF}, ClogP, and heavy atom count for each topology class are given in Table S2.

How the topological classes are related to molecular size and lipophilicity was also investigated. The results are shown in the Supporting Information. Figure S1 shows that the trend with respect to promiscuity and ClogP is the same for all the four topology classes and consistent with the overall behavior shown in Figure 3a. In contrast the relationship between promiscuity and molecular size differs significantly between the four topology classes and is generally irregular (Figure S2). Figure 8 shows that the promiscuity trends for the topology classes are more complex when the compounds are grouped according to their ClogP. For ClogP = 1−3 all the topological classes show roughly the same promiscuity, while for ClogP > 3 the 1TR class has the lowest promiscuity and the 3TR + B class the highest. The high median promiscuity for the 3TR + B class is especially noteworthy. Thus, for low f_{MF}, 1TR is the most common topological class and it is also generally the most selective one. For larger f_{MF} other topological classes are more common, and these are also in general more promiscuous. This trend is independent of lipophilicity for ClogP > 3. The results indicate that the ring system for compounds in the 1TR class makes significant contribution to the binding to the

primary target but is less likely to match secondary targets. The other topological classes might have larger conformational freedom that could result in higher promiscuity. In Figure 9 several representative examples of compounds from the BioPrint data set are displayed. The chosen molecules have a ClogP roughly in the interval between 2 and 3. In general, molecules with smaller f_{MF} containing only one ring system tend to be more selective. More promiscuous molecules usually have larger f_{MF} and more than one terminal ring system.

Conclusions

A correlation has been found between the descriptor f_{MF} and promiscuity for the BioPrint data set. Molecules with f_{MF} above 0.65 are generally more promiscuous than other molecules. Accordingly, molecules with a large molecular framework and only a few side chain atoms exhibit higher promiscuity than other types of molecules. Dividing the molecules into several topology classes showed that compounds with only one ring system were the most selective. It has also been shown that the effect of f_{MF} is not related to the lipophilic effect on promiscuity. Hydrophilic compounds with a large f_{MF} are still more promiscuous than hydrophilic compounds with a small f_{MF}. Molecules in the 1TR topological class show the highest selectivity. 1TR is also the most common topological class for molecules with an f_{MF} below 0.6. The topological classes 2TR + B and 3TR + B are more common and also more promiscuous for molecules with a large f_{MF}. The 3TR + B class exhibits especially high median promiscuity. In the future we are planning to investigate how f_{MF} relates to ADME properties as well as the relationship to other types of descriptors. In conclusion the results in this paper show that

the topology of the molecule is affecting the promiscuity for the BioPrint data set. The findings here indicate how selectivity can be modulated in drug discovery projects.

Supporting Information Available: Figure S1 showing the relationship between the median promiscuity and ClogP for each topology class; Figure S2 showing the relationship between median promiscuity and number of heavy atoms (Nheavy) for each topology class; Table S1 listing the number of compounds for each f_{MF} bin and topology class; Table S2 listing the median value of promiscuity, f_{MF}, and Nheavy for compounds in the four different topology classes. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Druker, B. J. STI571 (Gleevec) as a paradigm for cancer therapy. *Trends Mol. Med.* **2002**, *8*, S14–S18.

(2) Kalgutkar, A. S.; Crews, B. C.; Rowlinson, S. W.; Marnett, A. B.; Kozak, K. R.; Remmel, R. P.; Marnett, L. J. Biochemically based design of cyclooxygenase-2 (COX-2) inhibitors: facile conversion of nonsteroidal antiinflammatory drugs to potent and highly selective COX-2 inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 925–930.

(3) Mencher, S.; Wang, L. Promiscuous drugs compared to selective drugs (promiscuity can be a virtue). *BMC Clin. Pharmacol.* **2005**, *5*, 3.

(4) Stephenson, V. C.; Heyding, R. A.; Weaver, D. F. The "promiscuous drug concept" with applications to Alzheimer's disease. *FEBS Lett.* **2005**, *579*, 1338–1342.

(5) Roth, B. L.; Sheffler, D. J.; Kroeze, W. K. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat. Rev. Drug Discovery* **2004**, *3*, 353–359.

(6) Hampton, T. "Promiscuous" anticancer drugs that hit multiple targets may thwart resistance. *JAMA, J. Am. Med. Assoc.* **2004**, *292*, 419–422.

(7) Azzaoui, K.; Hamon, J.; Faller, B.; Whitebread, S.; Jacoby, E.; Bender, A.; Jenkins, J. L.; Urban, L. Modeling promiscuity based on in vitro safety pharmacology profiling data. *ChemMedChem* **2007**, *2*, 874–880.

(8) Rolland, C.; Gozalbes, R.; Nicolai, E.; Paugam, M.-F.; Coussy, L.; Barbosa, F.; Horvath, D.; Revah, F. G-Protein-coupled receptor affinity prediction based on the use of a profiling dataset: QSAR design, synthesis, and experimental validation. *J. Med. Chem.* **2005**, *48*, 6563–6574.

(9) Fliri, A. F.; Loging, W. T.; Thadeio, P. F.; Volkmann, R. A. Biological spectra analysis: linking biological activity profiles to molecular structure. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 261–266.

(10) Roche, O.; Schneider, P.; Zuegge, J.; Guba, W.; Kansy, M.; Alanine, A.; Bleicher, K.; Danel, F.; Gutknecht, E.-M.; Rogers-Evans, M.; Neidhart, W.; Stalder, H.; Dillon, M.; Sjogren, E.; Fotouhi, N.; Gillespie, P.; Goodnow, R.; Harris, W.; Jones, P.; Taniguchi, M.; Tsujii, S.; von der Saal, W.; Zimmermann, G.; Schneider, G. Development of a virtual screening method for identification of "frequent hitters" in compound libraries. *J. Med. Chem.* **2001**, *45*, 137–142.

(11) Hopkins, A. L.; Mason, J. S.; Overington, J. P. Can we rationally design promiscuous drugs? *Curr. Opin. Struct. Biol.* **2006**, *16*, 127–136.

(12) Kamal, A.; Jacques, H.; Bernard, F.; Steven, W.; Edgar, J.; Andreas, B.; Jeremy, L. J.; Laszlo, U. Modeling promiscuity based on in vitro safety pharmacology profiling data. *ChemMedChem* **2007**, *2*, 874–880.

(13) Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discovery* **2007**, *6*, 881–890.

(14) Whitlock, G. A.; Fish, P. V.; Fray, M. J.; Stobie, A.; Wakenhut, F. Pyridyl-phenyl ether monoamine reuptake inhibitors: impact of lipophilicity on dual SNRI pharmacology and off-target promiscuity. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2896–2899.

(15) Peters, J.-U.; Schnider, P.; Mattei, P.; Kansy, M. Pharmacological promiscuity: dependence on compound properties and target specificity in a set of recent roche compounds. *ChemMedChem* **2009**, *4*, 680–686.

(16) Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893.

(17) Bemis, G. W.; Murcko, M. A. Properties of known drugs. 2. Side chains. *J. Med. Chem.* **1999**, *42*, 5095–5099.

(18) Xu, Y.-J.; Johnson, M. Algorithm for naming molecular equivalence classes represented by labeled pseudographs. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 181–185.

(19) Xu, Y.-J.; Johnson, M. Using molecular equivalence numbers to visually explore structural features that distinguish chemical libraries. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 912–926.

(20) *Pipeline Pilot*, version 7.5; Accelrys: San Diego, CA.

(21) *ClogP*, version 4.3; BioByte Corp.: Claremont, CA.

(22) *OEChem*, version 1.3.4; OpenEye Scientific Software, Inc.: Santa Fe, NM.

(23) Krejsa, C. M.; Horvath, D.; Rogalski, S. L.; Penzotti, J. E.; Mao, B.; Barbosa, F.; Migeon, J. C. Predicting ADME properties and side effects: the BioPrint approach. *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 470–80.

(24) Wilcoxon, F. Individual comparisons by ranking methods. *Biom. Bull.* **1945**, *1*, 80–83.

(25) *JMP*, version 7; SAS Institute Inc.: Cary, NC.