Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

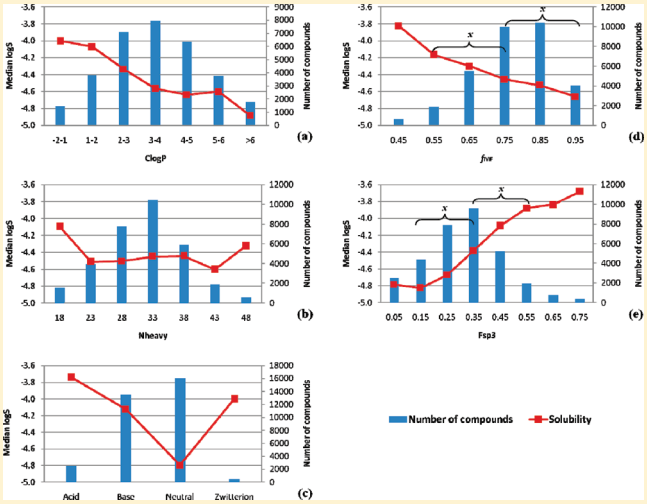
Beyond Size, Ionization State, and Lipophilicity: Influence of Molecular Topology on Absorption, Distribution, Metabolism, Excretion, and Toxicity for Druglike Compounds

Yidong Yang,*[,†,§] Ola Engkvist,[†] Antonio Llinàs,[‡] and Hongming Chen*[,†]

[†]Discovery Sciences, Computational Sciences, Computational Chemistry, and [‡]R&I iMED, *In Vitro* & *In Vivo* ADME, AstraZeneca R&D Mölndal, SE-431 83 Mölndal, Sweden

Supporting Information

ABSTRACT: The absorption, distribution, metabolism, excretion, and toxicity (ADMET) of a compound is dependent on physicochemical properties such as molecular size, lipophilicity, and ionization state. However, much less is known regarding the relationship between ADMET and the molecular topology. In this study two descriptors related to the molecular topology have been investigated, the fraction of the molecular framework (f_{MF}) and the fraction of sp^3-hybridized carbon atoms (Fsp^3). f_{MF} and Fsp^3, together with standard physicochemical properties (molecular size, ionization state, and lipophilicity), were analyzed for a set of ADMET assays. It is shown that aqueous solubility, Caco-2 permeability, plasma protein binding, human ether-a-go-go-related potassium channel protein inhibition, and CYP3A4 (CYP = cytochrome P450) inhibition are influenced by the molecular topology. These findings are in most cases independent of the already well-established relationships between the properties and molecular size, lipophilicity, and ionization state.



INTRODUCTION

For many years research-based pharmaceutical companies have faced considerable difficulties with high attrition rates for compounds entering clinical development. Historically, one of the main reasons for failure was poor pharmacokinetics (PK) and bioavailability.[1,2] Since the 1990s, distribution, metabolism, and pharmacokinetics (DMPK) have been addressed already in the lead generation phase, which has led to a decrease in clinical attrition due to DMPK. However, efficacy and toxicity have instead become major contributors to the overall compound related attrition.[3,4]

In the period 1992−2002, 33% of the candidate drugs in clinical phases I−III were terminated due to toxicity and over 90% of the drug withdrawals from the market were caused by toxicity.[3] The average preapproval cost for a new drug is nowadays estimated to exceed $800 million.[5] These facts highlight the potential risk and accompanying costs associated with not identifying toxic side effects of a promising drug candidate until in the late clinical development or even in the postapproval phase. As a result, safety screens, such as inhibition of human cytochrome P450 (CYP) enzymes and the human ether-a-go-go-related gene (hERG) ion channel, are needed as part of the regulatory requirements.

To improve the chances of success for a candidate drug in development, current drug discovery programs have adopted a strategy to investigate absorption, distribution, metabolism, excretion, and toxicity (ADMET) early and in parallel with structure−activity relationship (SAR) studies during lead generation and optimization.

Since Lipinski et al.[6] profiled a range of physicochemical properties of drugs and derived the well-known "rule of 5", an increasing body of literature suggests that poor ADMET outcomes are predominantly correlated with increasing lipophilicity and molecular size.[7−12] It has been shown that compounds in later clinical trial stages have lower lipophilicity than those in phase I.[10] Lipophilic compounds have also been associated with promiscuity.[8] Thus, the understanding of the importance to control ADMET by optimizing key physicochemical properties, such as lipophilicity, has increased during the past decade.[13] With this relationship firmly established, the attention has been changed to find other molecular descriptors that can be correlated to clinical success through their influence on ADMET.

Recent studies provide additional molecular descriptors to complement the well-established physicochemical properties, size, lipophilicity, and ionization state: i.e., aromatic ring count,[14,15] fraction of sp^3-hybridized carbons (Fsp3),[16,17] chiral atom

Received: July 1, 2011
Published: March 26, 2012

counts,[16] aromatic atom count − sp^3 atom count,[18] and how the fraction of the molecular framework (f_{MF}) is related to promiscuity.[19] According to observations based on the aromatic ring count, chiral carbon atom count, and Fsp^3, it is suggested that a more three-dimensional structure is associated with more favorable druglike attributes.[20] The correlation between f_{MF} and promiscuity[19] indicates that a smaller molecular framework and more side chain atoms will improve selectivity. The effect of f_{MF} is related to other descriptors such as the number of rotatable bonds. However, while f_{MF} is straightforward to calculate, the number of rotatable bonds might be defined in many different ways, which might affect the interpretation of experimental data.[21] Notably, the effect of f_{MF} on promiscuity is not related to the size and lipophilicity. However, it is clear that the aromatic ring count and chiral atom count are size-dependent properties. It has also been shown that the aromatic ring count correlates positively with lipophilicity as measured by ClogP and log *D*.[13]

Table 1. Correlation Coefficients (r^2) between the Descriptors Used in This Study

descriptor	ClogP	Nheavy	f_{MF}	Fsp^3
ClogP		0.06	0.005	0.003
Nheavy	0.06		0.001	0.06
f_{MF}	0.005	0.001		0.016
Fsp^3	0.003	0.06	0.016	

In the current study it is first shown that f_{MF} and Fsp^3 are independent of the molecular size and ClogP, followed by an investigation of how f_{MF} and Fsp^3 are related to important

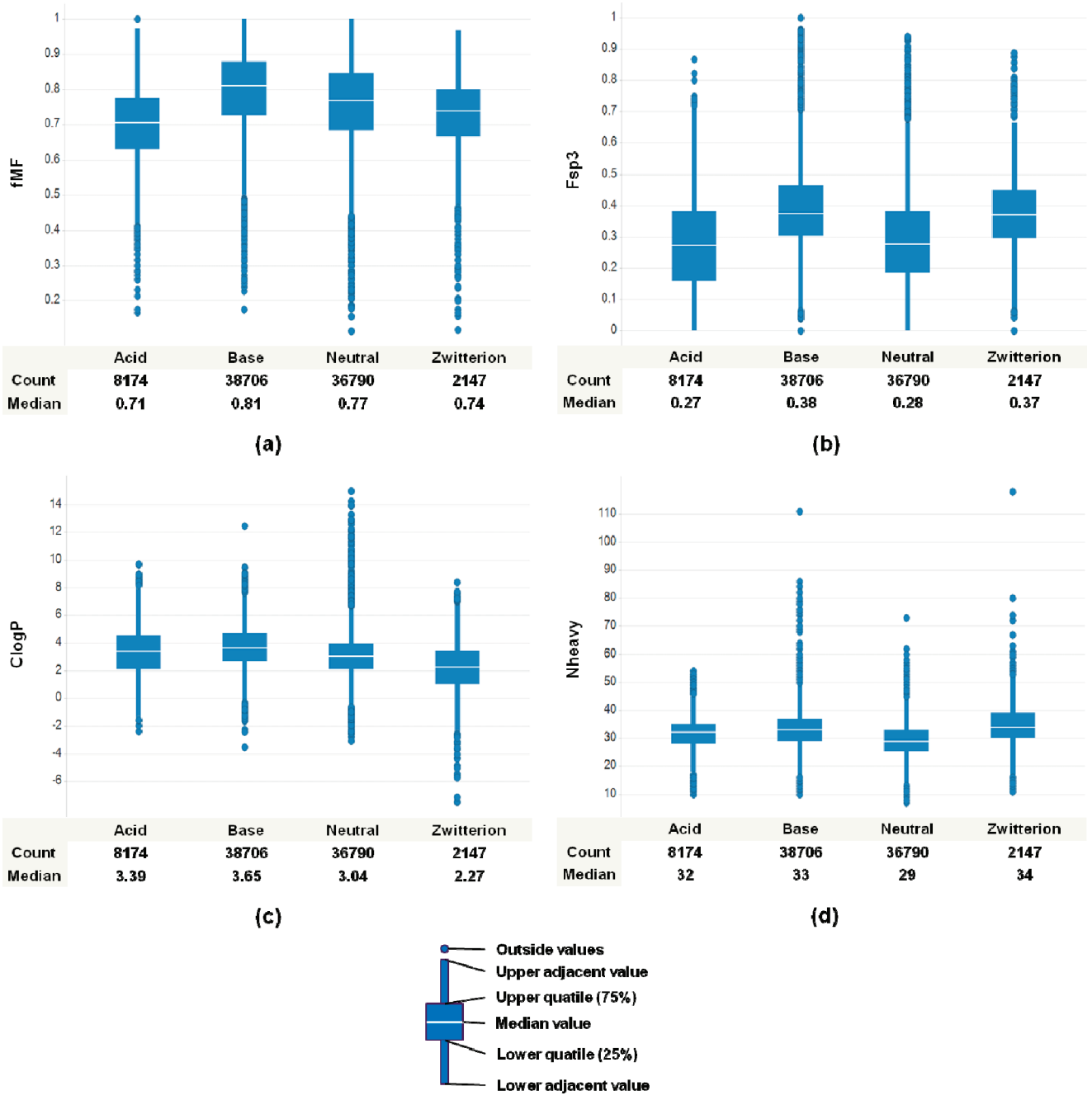


Figure 1. Comparison of the median f_{MF} (a), Fsp^3 (b), ClogP (c), and Nheavy (d) for different ionization states.

dx.doi.org/10.1021/jm201548z | *J. Med. Chem.* 2012, 55, 3667−3677

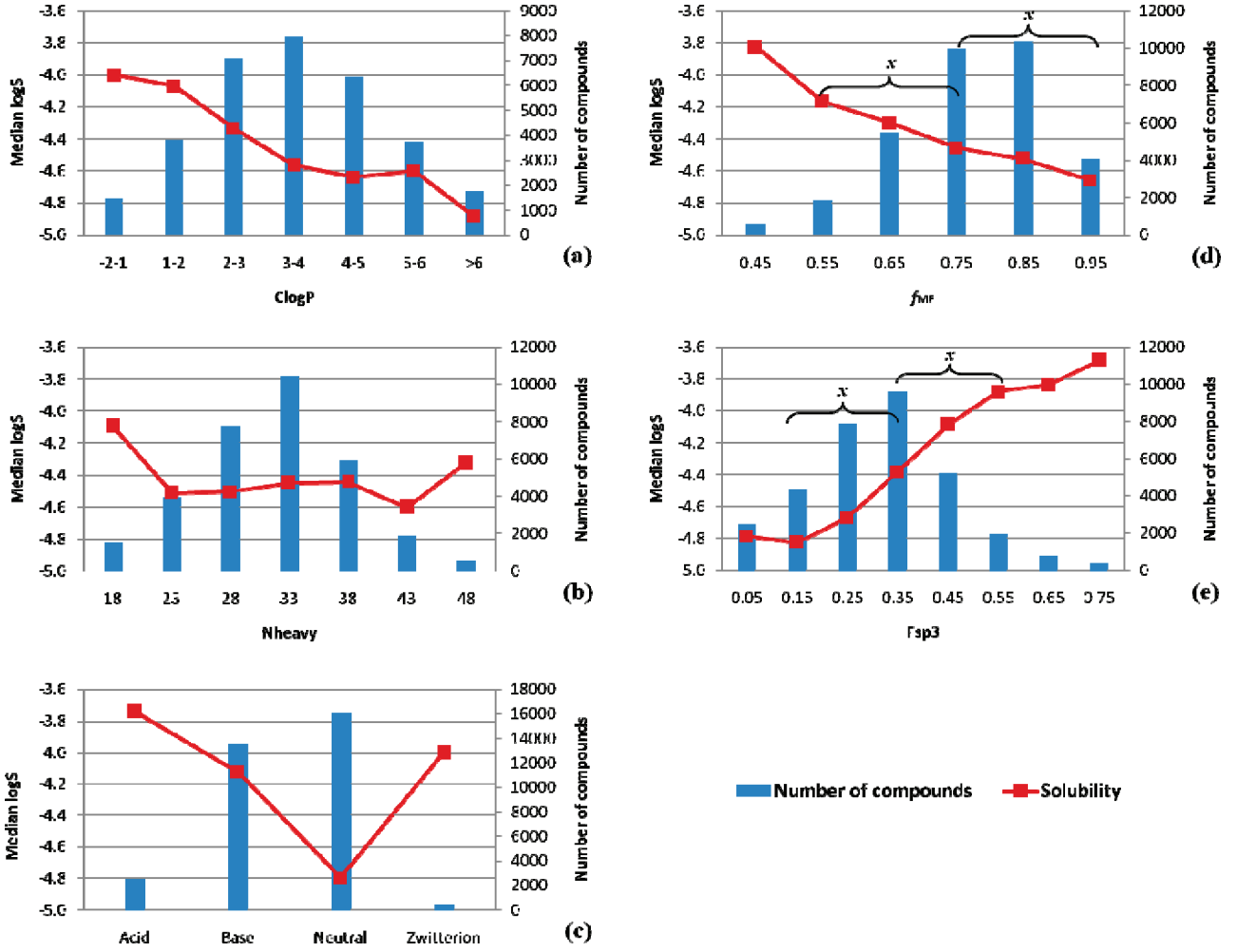


Figure 2. Relationship between the aqueous solubility (log *S*, μM) and ClogP (a), Nheavy (b), ionization state (c), f_{MF} (d), and Fsp^3 (e) [(*x*) $p < 0.0001$ (*p* is the probability that the distributions for the bins are indistinguishable according to the Wilcoxon rank-sum test)].

ADMET assays, such as aqueous solubility, permeability, plasma protein binding (PPB), and hERG and CYP3A4 inhibition. Hopefully, this investigation together with those from other groups will provide more extensive guidelines for selecting and prioritizing compounds beyond just reducing the lipophilicity and molecular size. An increased understanding of the relationship between molecular topology and druglike properties is highly desirable in library design and profiling. Additional criteria based on the molecular topology would complement existing criteria related to physicochemical properties. To our knowledge, only one large study of the influence of size, lipophilicity, and ionization state on ADMET has been reported in the literature.[7] Even though this study investigates the relationship between topological descriptors and ADMET, it is also of general interest to verify earlier conclusions by an independent data set, to further establish the relationship between experimentally measured ADMET data and physicochemical descriptors.

RESULTS AND DISCUSSION

The physicochemical descriptors selected in this study are similar to the ones used previously. Gleeson et al.[7] selected, after a principal component analysis (PCA), size (molecular weight), ClogP, and ionization state as descriptors for analyzing ADMET data. Similarly, we used as a complement to the topological descriptors f_{MF} and Fsp^3 the number of heavy atoms (Nheavy) as a measure of molecular size, ClogP as a measure of lipophilicity, and the ionization state.

First, the intercorrelation among ClogP, Nheavy, f_{MF}, and Fsp^3 was investigated. As seen in Table 1 the four descriptors are not correlated with each other for the whole data set. The correlation matrix is based on 86 115 compounds tested in the ADMET assays and used in this analysis. The results indicate that the impact of the topological descriptors on ADMET (demonstrated individually in the following subsections) is effectively independent of the molecular size and lipophilicity. However, the independence needs to be reconfirmed for each specific ADMET assay, since it might be possible that there exists a correlation for the compounds measured in that particular assay.

For each ADMET assay, compounds were binned and the bins containing less than 1% of the whole data set were removed. The bins were of equal size, leading to different numbers of molecules in each bin. It was felt that this binning scheme was the most relevant for this investigation. However, to prove the validity of our findings, the calculations were repeated with a binning scheme that populates each bin with an equal number of molecules. The results for the most important identified correlations with this alternative binning scheme are included in the Supporting Information. The Supporting Information also includes information regarding standard

dx.doi.org/10.1021/jm201548z | J. Med. Chem. 2012, 55, 3667−3677

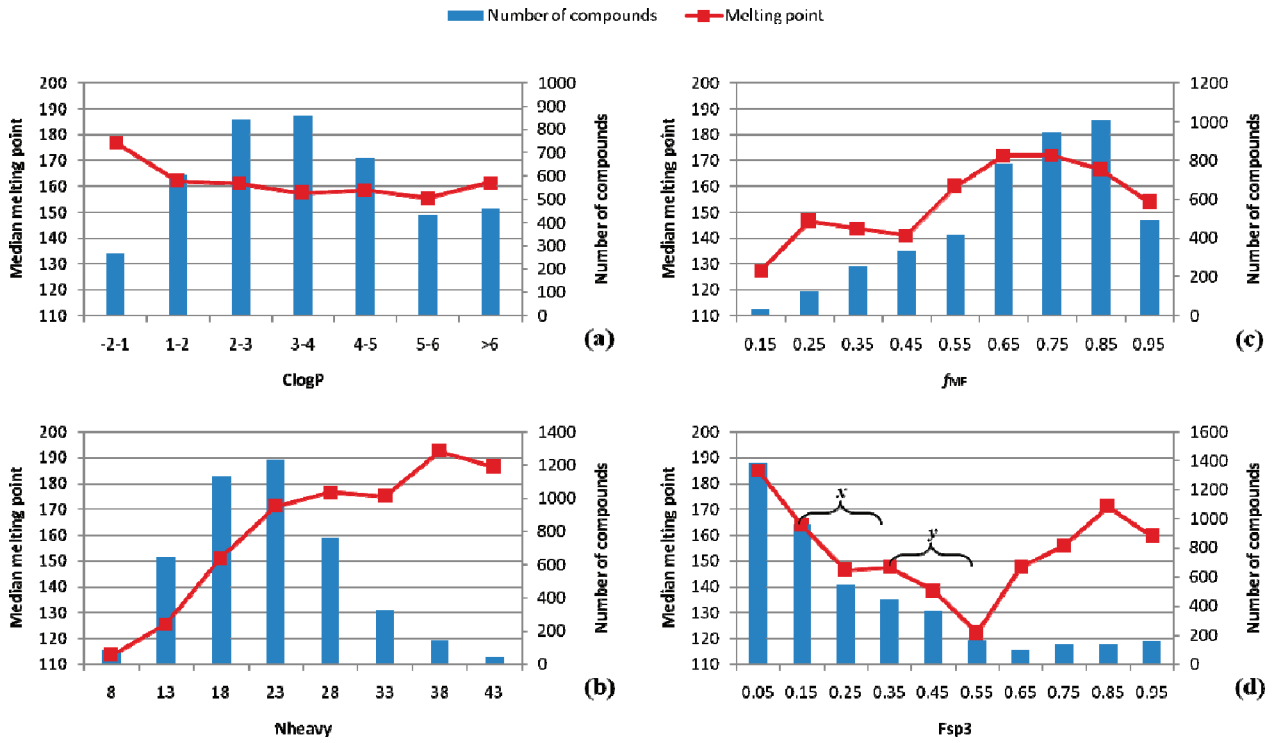


Figure 3. Relationship between the melting point and ClogP (a), Nheavy (b), f_{MF} (c), and Fsp^3 (d) [(*x*) $p < 0.0001$; (*y*) $p < 0.005$].

deviations for the binned descriptors. For results that are extensively discussed in the text, we have included statistical tests (Wilcoxon rank-sum) in the relevant figures. Differences between other bins might also be statistically significant. However, since the differences are not explicitly discussed in the text, the corresponding statistical tests are not included in the figures. The median values for f_{MF}, Fsp^3, ClogP and Nheavy, calculated for the whole data set and partitioned according to the ionization state, are compared in Figure 1. Basic compounds have a higher median for f_{MF}, Fsp^3, and ClogP than acidic and neutral compounds.

Aqueous Solubility and Melting Point. For a drug to be absorbed, it has to be soluble and permeable. High solubility provides an essential free concentration of the compound that can permeate across a biological membrane. Low aqueous solubility can result in low absorption, even if the permeability is good. Low solubility has therefore been identified as the cause of numerous drug development failures.[22] The strong correlation between lipophilicity and low aqueous solubility has been extensively described.[23−25]

As expected, an analysis of over 32 000 in-house compounds shows that the median aqueous solubility decreases with increasing ClogP (Figure 2a), which is consistent with earlier studies.[7] The relation between molecular size and aqueous solubility is fairly constant, except for the smallest and largest molecules, which have higher solubility (Figure 2b). The relationship between solubility and ionization state is shown in Figure 2c, and again, as expected, charged molecules have higher solubility than neutral molecules at physiological pH, with acidic compounds having the highest solubility. Roughly the same trend has been observed previously; however, in that study zwitterionic compounds had the highest solubility.[7]

Figure 2d shows a strong relationship between solubility and f_{MF}. Aqueous solubility decreases significantly for large f_{MF}. There is also a strong correlation between Fsp^3 and solubility. An increase of Fsp^3 is correlated with higher solubility[16] (Figure 2e). The trends for f_{MF} and Fsp^3 are the same for all ionization states (Figure S1, Supporting Information). The only exception is for acidic compounds and Fsp^3, where for large values of Fsp^3 the solubility decreases. However, it should be noted that for these high values for Fsp^3 the measured solubility data are sparse.

It has been shown previously[26−28] that the solubility of an organic compound is closely related to its melting point, since the melting point of a compound reflects the crystal lattice energies. Breaking up the crystal lattice is the first step in the dissolution process. The relationship between the melting point and the descriptors was investigated to better understand the influence of the descriptors on the solubility. The used melting point data set is freely available and consists of 4445 compounds.[29]

The melting point is fairly constant for the different ClogP intervals, except for the most hydrophilic compounds, which have higher median melting points (Figure 3a). The melting point increases significantly with increasing size (Figure 3b). Figure 3c shows that compounds with larger f_{MF} have significantly higher median melting points than those with smaller f_{MF}. However, in the range above 0.5, the difference in the median melting point is in the range of 15 deg, peaking at an f_{MF} of approximately 0.7. Thus, for the f_{MF} range above 0.50, which is the most relevant range for the compounds measured in the solubility assay, there is no strong correlation between the melting point and f_{MF}. This might indicate that the relationship between f_{MF} and aqueous solubility is not due to strong intermolecular interactions in the crystal. However, it should be emphasized that the data sets for the solubility and melting point measurements are different.

As Fsp^3 increases, the median melting point in the lower range of Fsp^3 decreases (Figure 3d). This is in agreement with earlier results.[16] The correlation between melting point and Fsp^3 might be due to the fact that molecules with a lower Fsp^3

dx.doi.org/10.1021/jm201548z | J. Med. Chem. 2012, 55, 3667−3677

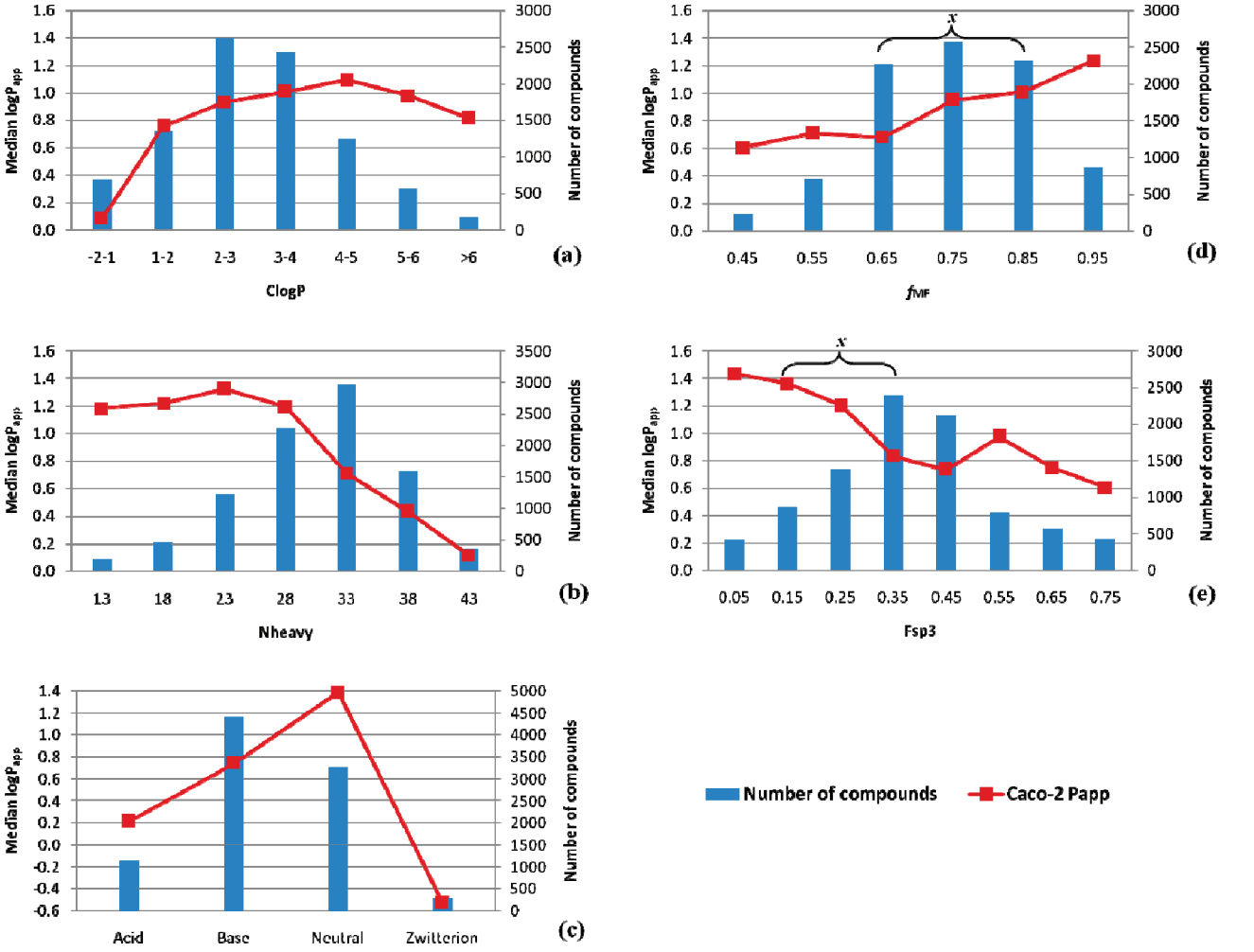


Figure 4. Relationship between the apparent permeability (log P_{app}, nm/s) and ClogP (a), Nheavy (b), ionization state (c), f_{MF} (d), and Fsp^3 (e) [(*x*) $p < 0.0001$].

are more easily involved in π-stacking, forming stronger crystal interactions. However, other factors such as symmetry of the molecules can also influence the melting point and accordingly the solubility. As Fsp^3 increases, the molecules tend to have lower melting points, possibly due to disruption of the planarity and symmetry.[20] The melting point starts to increase for Fsp^3 values above 0.55. However, since very few druglike compounds have such a high value for Fsp^3, it is not clear if this trend is of practical interest. Thus, the external melting point data set clearly shows that the melting point decreases with increasing Fsp^3, which indicates that the correlation observed between solubility and Fsp^3 for the in-house data set is at least in part due to intermolecular interactions in the crystal. The relationships between f_{MF} and Fsp^3 and the aqueous solubility is still valid for an alternative binning scheme.[30]

It is also important to investigate if f_{MF} and Fsp^3 are correlated with lipophilicity for this particular data set, even though, as shown in Table 1, there is no overall correlation between the descriptors and lipophilicity. Figure S2 (Supporting Information) shows a correlation between f_{MF} and ClogP; however, there is no correlation between Fsp^3 and ClogP. Since solubility is a monotonic function of ClogP, f_{MF}, and Fsp^3, it is possible to estimate their respective contributions to solubility from Figure 2. Fsp^3 and ClogP influence solubility to the same degree, while f_{MF} has a smaller influence. In conclusion, since Fsp^3 influences solubility as much as ClogP, it is an important descriptor to take into account when the aqueous solubility needs to be improved in lead generation and optimization.

Caco-2 Permeability. To be bioavailable, an orally administered drug needs to cross the intestinal epithelium to reach the blood circulation. There are different mechanisms to pass the membrane, including transcellular diffusion, paracellular diffusion, active carrier-mediated transport, and transcytosis. Caco-2 cells are commonly used to predict the intestinal permeability, since they express a number of transporters. They have tight junctions, giving the cells a functionality similar that of the cells lining the small intestine.[31]

The median permeability correlates with ClogP in the lower range (Figure 4a). A peak for the permeability is observed at a ClogP between 4 and 5. The permeability decreases then slightly for higher values of ClogP. It is logical that the permeability first increases with increasing ClogP, since the membrane is hydrophobic and the desolvation energy is smaller for more lipophilic compounds. That the permeability slightly decreases for high ClogP might be due to the compounds remaining in the membrane instead of permeating through. To cross a membrane, a molecule has to partition from the aqueous phase into the hydrophobic layer (rate-limiting step for hydrophilic compounds) and again partition into the

dx.doi.org/10.1021/jm201548z | J. Med. Chem. 2012, 55, 3667−3677

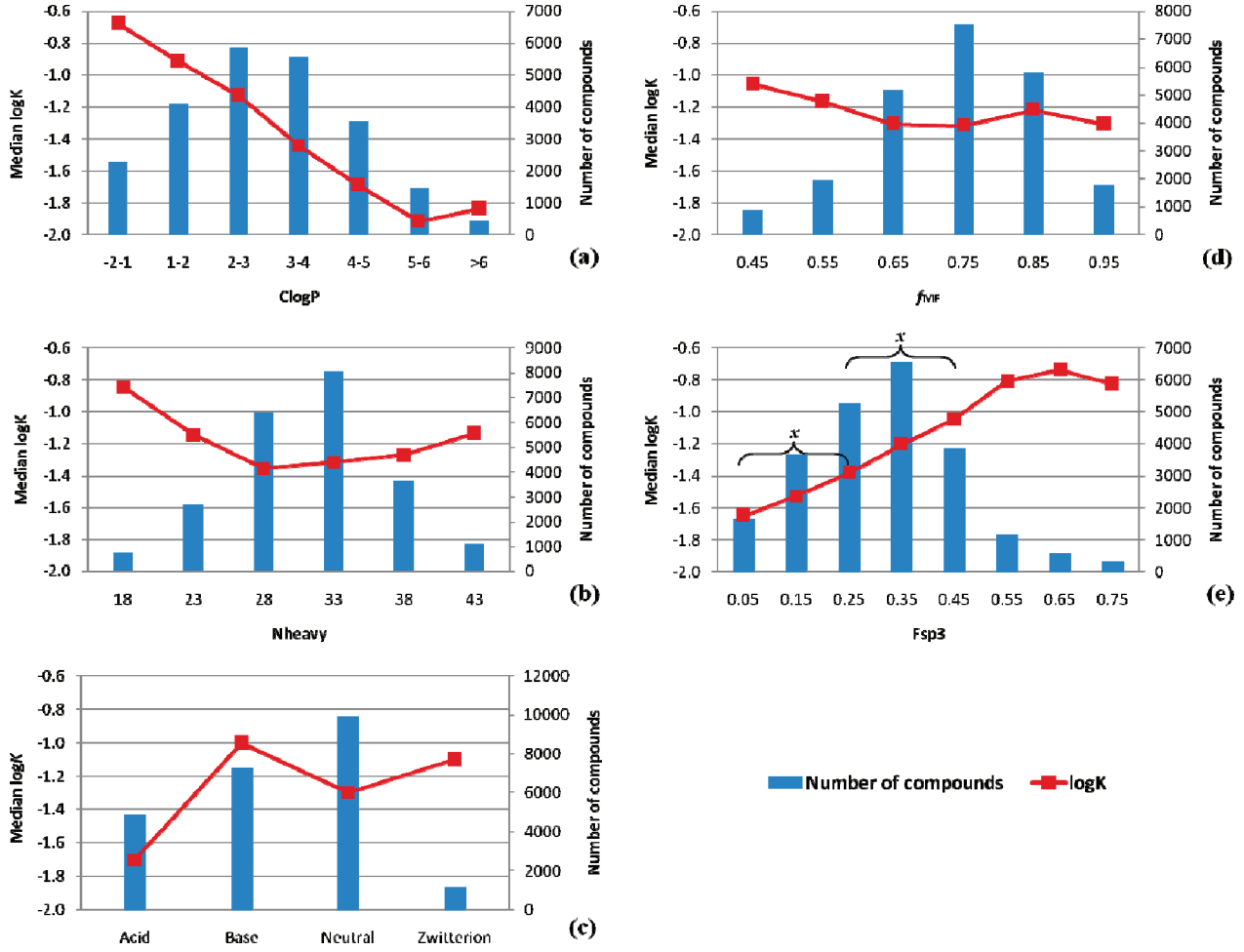


Figure 5. Relationship between plasma protein binding (log K) and ClogP (a), Nheavy (b), ionization state (c), f_{MF} (d), and Fsp^3 (e) [(x) $p < 0.0001$].

aqueous phase when exiting the membrane (rate-limiting step for hydrophobic compounds).

The same trend as in Figure 4 has been observed for artificial membrane permeability of neutral compounds. However, in this experimental setup the permeability for ionized molecules increases monotonically with increasing ClogP.[7] As expected, permeability decreases as the molecule increases in size (Figure 4b); this is in agreement with other studies.[6,7,32−34] This effect is related to the polar surface area (PSA). Since the PSA is proportional to the size, the desolvation energy needed for permeation increases with increasing PSA. Neutral compounds have higher permeability than charged compounds due to larger desolvation energies for ionized compounds in general and for acids in particular (Figure 4c).

As f_{MF} increases, the median permeability also increases (Figure 4d). The trend remains when the data set is partitioned into different ionization states, though the trend is significantly weaker (Figure S3, Supporting Information). Fsp^3 displays an inverse relationship with permeability, in particular for compounds with an Fsp^3 lower than 0.35 (Figure 4e). There is no correlation between Fsp^3 and ClogP for the investigated data set, which emphasizes the independent role of Fsp^3 for permeability (Figure S4, Supporting Information). The permeability decreases with increasing Fsp^3 for acids and neutral compounds (Figure S3). Thus, both f_{MF} and Fsp^3 influence permeability. However, the trends are weaker when taking into account the ionization state.

Plasma Protein Binding. When a compound enters the bloodstream, it can be either bound to a plasma protein or unbound. The ability of plasma proteins to carry and distribute compounds is an important property for understanding the pharmacokinetic (PK) and pharmacodynamic (PD) properties of a compound. fu is needed to estimate the human dose. However, it is usually a mistake to optimize the chemical structure to reduce the PPB (decrease fu) and expect to see an increased in vivo efficacy due to a higher free drug plasma concentration. The average free drug concentration in vivo after oral dosing is, in most cases, independent of PPB.[35,36] In this study we have used log K, (log(fu/(100 − fu)), which is the ratio between a compound's unbound fraction and bound fraction, to represent the plasma protein binding.[37]

An analysis of 23 228 in-house compounds with measured log K shows a linear decrease with ClogP, which is also consistent with earlier studies[38−42] (Figure 5a). As can be seen in Figure 5b, the relationship between log K and molecular size is fairly constant for normally sized compounds. The trend for log K is basic > zwitterionic > neutral > acidic compounds (Figure 5c). This is generally in line with earlier results.[7]

The median log K is rather independent of f_{MF} (Figure 5d). However, partioning the data into different ionization states shows that log K decreases with increasing f_{MF} for basic, neutral, and zwitterionic compounds, while the behavior is irregular for acids (Figure S5, Supporting Information). There is a clear

dx.doi.org/10.1021/jm201548z | J. Med. Chem. 2012, 55, 3667−3677

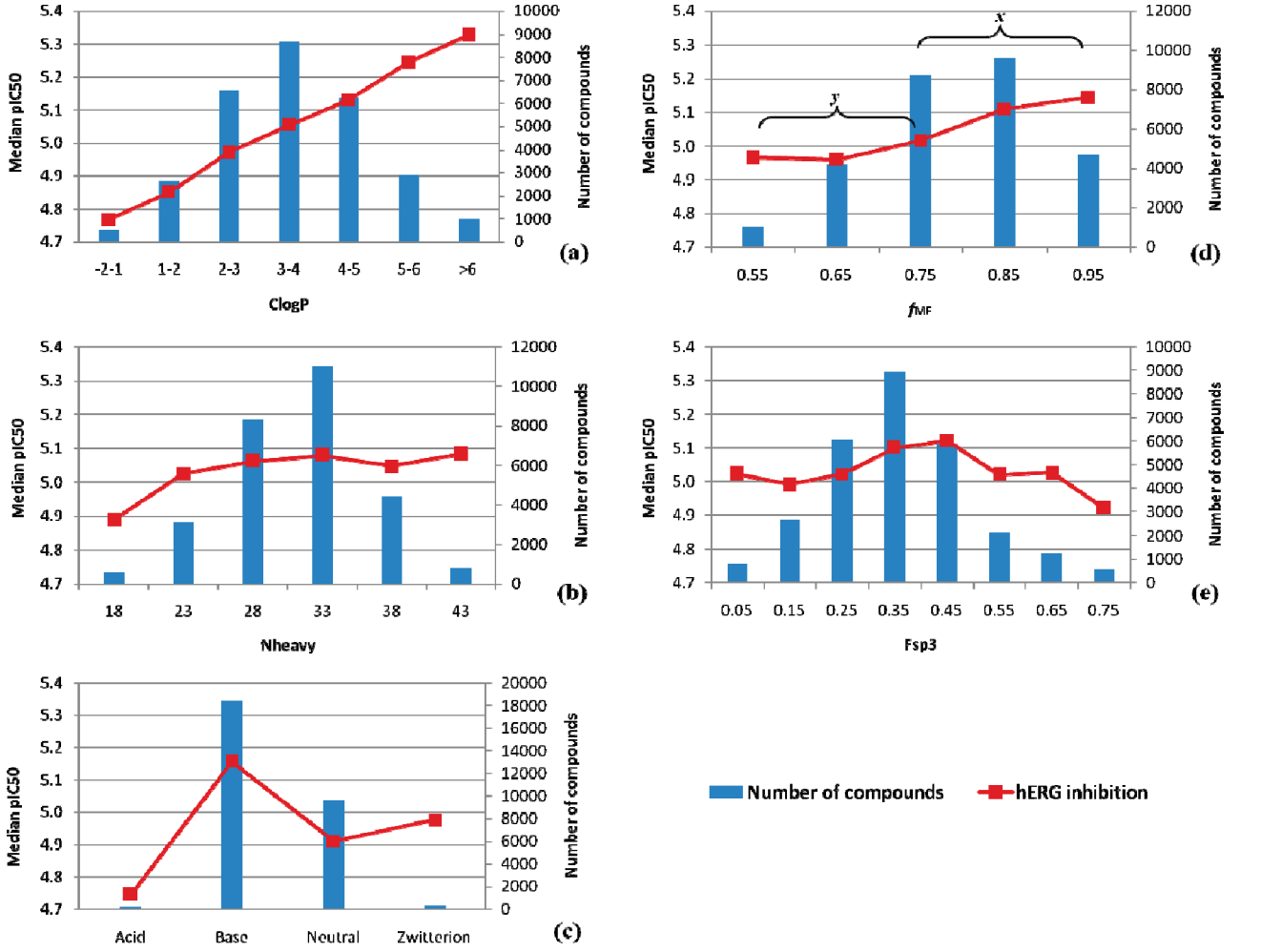


Figure 6. Relationship between hERG inhibition (pIC_{50}) and ClogP (a), Nheavy (b), ionization state (c), f_{MF} (d), and Fsp^3 (e) [(*x*) $p < 0.0001$; (*y*) $p < 0.0003$].

correlation between log *K* and Fsp^3 (Figure 5e). log *K* increases with increasing Fsp^3. This result is still valid after the data set is partitioned into ionization states (Figure S6, Supporting Information). While Fsp^3 has a large influence on log *K*, it is shown in Figure 5 that the influence of ClogP is larger. Fsp^3 is relatively independent of ClogP (Figure S7, Supporting Information), highlighting that Fsp^3 is an important descriptor for log *K*. The results described above are still valid with an alternative binning scheme for the data (Figure S12f, Supporting Information).

hERG Inhibition. Inhibition or compromising the function of the voltage-gated potassium ion channel, a protein encoded by the hERG, can result in a potential fatal disorder, long QT syndrome. Prolongation of the QT interval can lead to Torsades de pointes, a condition associated with a fall in arterial blood pressure, fainting, ventricular fibrillation, and sudden death.[43]

hERG inhibition was measured in an hERG IonWorks assay. From an analysis of 28 533 compounds, the median pIC_{50} displays a linear increase with ClogP (Figure 6a), which is in line with results reported in the literature.[44−46] Figure 6b shows that the median hERG pIC_{50} is not correlated with the molecular size.

Figure 6d shows that compounds with a larger f_{MF} have higher hERG inhibition. However, the change is not as pronounced as for ClogP. The relationship between Fsp^3 and hERG inhibition is not monotonic; compounds with an Fsp^3 of 0.45 have the highest median hERG inhibition, and compounds with an Fsp^3 of 0.75 have the lowest. As expected, basic compounds have the highest median hERG inhibition (Figure 6c). f_{MF} is monotonically correlated with hERG inhibition for positive and neutral compounds (Figure S8, Supporting Information). In conclusion, the largest influence on hERG inhibition is from ClogP and the ionization state; however, the molecular topology also influences the hERG inhibition as is shown for f_{MF}. The effect of f_{MF} on hERG inhibition is not related to lipohilicity (Figure S9, Supporting Information). The results are still valid with an alternative binning scheme for the data (Figure S12i, Supporting Information). Compounds with a large f_{MF} are more promiscuous,[19] and it is therefore not surprising that they also have higher affinity for the hERG ion channel. However, Figure 6 shows that the influence of ClogP on hERG inhibition is larger than the influence of f_{MF}.

CYP3A4 Inhibition. CYP3A4 is a gene encoding a member of the cytochrome P450 superfamily of enzymes, which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids, and other lipids. CYP3A4 is chosen here because it is the most abundant cytochrome P450 isoform[47] and it is involved in the metabolism of approximately half of the drugs currently in use.[48]

An analysis of 15 888 in-house-measured compounds showed that CYP3A4 inhibition is positively correlated with ClogP

dx.doi.org/10.1021/jm201548z | J. Med. Chem. 2012, 55, 3667−3677

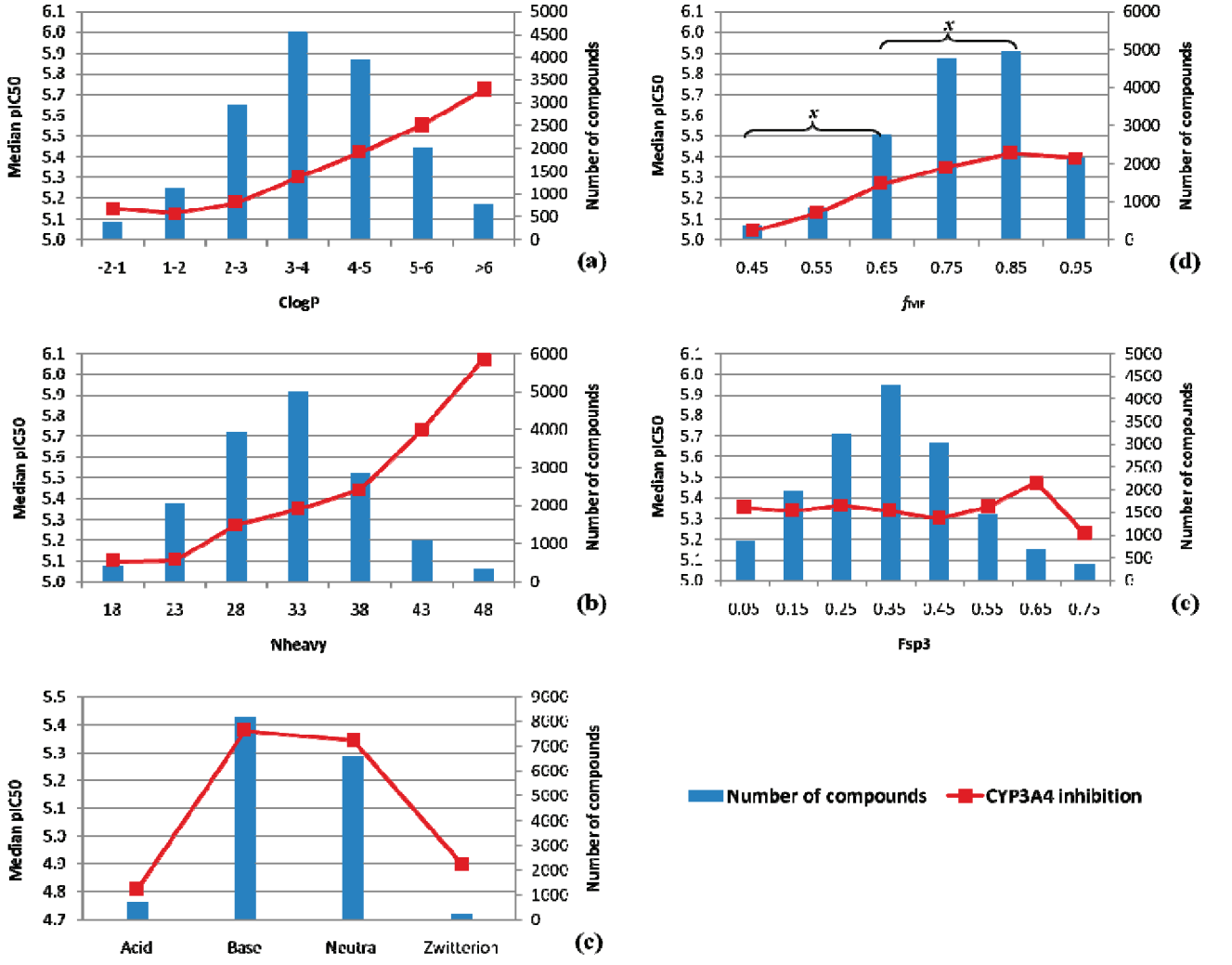


Figure 7. Relationship between CYP3A4 inhibition (pIC_{50}) and ClogP (a), Nheavy (b), ionization state (c), f_{MF} (d), and Fsp^3 (e) [(*x*) $p < 0.0001$].

Table 2. Summary of the Influence of f_{MF} and Fsp^3 on the Investigated ADMET Assays[a]

	Fsp3	f_{MF}
aqueous solubility	Fsp3 ↑, solubility ↑	f_{MF} ↑, solubility ↓
Caco-2 permeability	Fsp3 ↑, Caco-2 ↓	f_{MF} ↑, Caco-2 ↑
hPPB	Fsp3 ↑, fu ↑	f_{MF} ↑, fu ↓
hERG inhibition	no influence	f_{MF} ↑, hERG inhibition ↑
CYP3A4 inhibition	no influence	f_{MF} ↑, CYP3A4 inhibition ↑

[a]An upward arrow means increasing and a downward arrow decreasing for the descriptors and properties.

(Figure 7a). The median CYP3A4 inhibition increases, as expected, with increasing molecular size (Figure 7b). Positively charged and neutral compounds display higher CYP3A4 inhibition than negatively charged compounds and zwitterions (Figure 7c). CYP3A4 inhibition is also influenced by f_{MF} (Figure 7d). Compounds with a larger f_{MF} have higher CYP3A4 inhibition. The trend is valid for all three major ionization states (Figure S10, Supporting Information). Fsp^3 does not have an effect on CYP3A4 inhibition (Figure 7e). The influence of f_{MF} on CYP3A4 inhibition is smaller than the influence of ClogP and the ionization state; however, it is still significant. The effect of f_{MF} on CYP3A4 inhibition is unrelated to the effect of lipophilicity and size (Figure S11, Supporting Information). The results are still valid with an alternative binning scheme (Figure S12g,h, Supporting Information). For cases in lead optimization where it is difficult to decrease ClogP and molecular size, it might be a viable alternative to lower the f_{MF} to avoid CYP3A4 inhibition. As already discussed for hERG inhibition, compounds with a large f_{MF} are more promiscuous, so it is not surprising that compounds with a large f_{MF} also have higher CYP3A4 inhibition. However, Figure 7 shows that the influence of f_{MF} on CYP3A4 inhibition is smaller than the influence of lipophilicity and size.

CONCLUSIONS

The purpose of this study was to investigate how several important ADMET properties are influenced by two molecular descriptors that are related to the molecular topology, f_{MF} and Fsp^3. The descriptors are uncorrelated with both the molecular size and lipophilicity. They are both easy to calculate, and it should be straightforward to reproduce our descriptor calculations by others. The relationships between the descriptors and five important ADMET properties were investigated.

From the results reported herein, not only was it confirmed that molecular size, lipophilicity, and ionization state are very important descriptors for ADMET, but it was also shown that ADMET is influenced by the molecular topology. Both f_{MF} and Fsp^3 influence the aqueous solubility. The solubility decreases

dx.doi.org/10.1021/jm201548z | J. Med. Chem. 2012, 55, 3667−3677

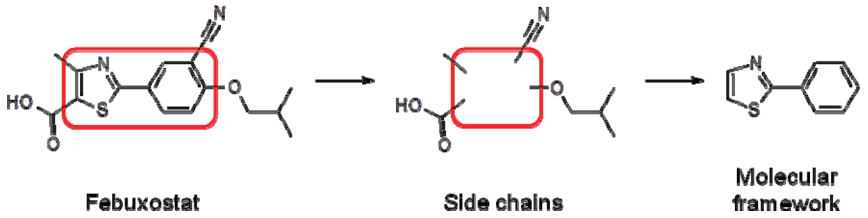


Figure 8. Disconnecting the side chains from the original molecule defines its MF. f_{MF} of febuxostat can be calculated as the number of heavy atoms in the MF (11) divided by the total number of heavy atoms in the molecule (22). Accordingly, f_{MF} is 0.5. The number of sp^3 carbons is 5, and the total number of carbon atoms is 16; thus, Fsp^3 is 0.31.

with increasing f_{MF} and increases with increasing Fsp^3. These trends are independent of the ionization state. The melting point is dependent on Fsp^3, indicating that the origin of the influence for Fsp^3 on the aqueous solubility is at least partly due to the crystal lattice energy. Caco-2 absorption is also influenced by f_{MF} and Fsp^3. However, the influence is rather small when the ionization state is taken into account. Plasma protein binding is strongly influenced by Fsp^3; increasing Fsp^3 decreases the plasma protein binding. hERG inhibition and CYP inhibition are influenced by f_{MF}, while independent of Fsp^3. The results are summarized in Table 2.

It is noted that most results found in this study are empirical; there is currently no explanation why for instance plasma protein binding shows a strong correlation with Fsp^3. It should also be noted that even though the topological descriptors Fsp^3 and f_{MF} are usually independent of size and lipophilicity, they still show a dependency on the ionization state. It is crucial to partition the experimental ADMET data according to the ionization state when analyzing the influence of Fsp^3 and f_{MF}. Additionally, this study confirms earlier results relating physicochemical properties to ADMET.[7]

Given the fact that the two topological descriptors can easily be calculated and interpreted, it may provide an opportunity for medicinal chemists to modify the structures to achieve optimal ADMET. Especially, in cases where lowering the lipophilicity is not an option due to the need for potency on the primary target, modifying the molecular topology might be a viable option to improve the ADMET of a molecule. We are currently investigating how the identified relationships between topology and ADMET might influence library design, library profiling, and lead optimization.

EXPERIMENTAL SECTION

Experimental Data. Experimental data from the different ADMET assays were extracted from an in-house database. These include assay data for aqueous solubility, Caco-2 permeability, human plasma protein binding (fu), hERG inhibition, and CYP3A4 inhibition. The aqueous solubility data set consists of 32 549 compounds measured in a phosphate buffer at pH 7.4 at room temperature (22 °C) after 18−24 h of equilibrium.[49] The Caco-2 permeability data set consists of 9107 compounds.[50] Plasma protein binding was determined using equilibrium dialysis for 23 228 compounds.[51] A set of 28 533 compounds measured in an hERG IonWorks electrophysiology assay[52] and CYP3A4 inhibition data[53] for 15 888 in-house compounds were used in the analysis. In addition to the in-house data, the melting point data for 4445 compounds reported by Karthikeyan et al.[29] were downloaded from http://cheminformatics.org/.

Molecular Descriptors. ClogP was calculated with a commercial program.[54] The descriptor f_{MF} is defined as the number of heavy atoms in the molecular framework (MF) divided by the total number of heavy atoms (Figure 8).[19] Fsp^3 is defined as the number of sp^3-hybridized carbons divided by the total number of carbon atoms (Figure 8).[16] These two descriptors were generated with Pipeline Pilot.[55] f_{MF} is strictly defined by the molecular topology and should therefore be completely independent of which program has been used for the calculation. Nevertheless, Fsp^3 might be dependent on the program used since different programs might define aromaticity differently. However, for druglike molecules such as the ones analyzed in this study, it is in most cases straightforward to determine which rings are aromatic. It is therefore not expected that different programs will give different values for Fsp^3 for a significant number of druglike molecules. The ionization state was determined by substructure matching of a set of predefined acidic, basic, and cationic functional groups with an in-house-developed program.

The nonparametric Wilcoxon rank-sum test[56] was applied to determine whether the differences between two distributions were statistically significant. All statistical analyses were performed with JMP.[57]

ASSOCIATED CONTENT

Supporting Information

Figures S1−S12 and an Excel sheet which includes the standard deviation for each bin in Figures 2−7 and S1−S12. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: +46 (0)31 7065285 (H.C.); +46 (0)31 7061097 (Y.Y.). Fax: +46 (0)31 7763792 (Y.Y.); +46 (0)31 7763792 (H.C.). E-mail: hongming.chen@astrazeneca.com (H.C.); yidong@gmail.com (Y.Y.).

Present Address

[§]Crown Bioscience Inc., Taicang, Jiangsu Province, People's Republic of China.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Niklas Blomberg, Dr. Barry Collins, Dr. Ken Grime, and Dr. Paul Leeson for valuable discussions and comments on the manuscript.

ABBREVIATIONS USED

MF, molecular framework; f_{MF}, fraction of the MF, number of atoms in the MF divided by the total number of atoms in the molecule; Fsp^3, fraction of sp^3-hybridized carbons; Nheavy, number of heavy atoms; PK, pharmacokinetics; DMPK, drug metabolism and pharmacokinetics; CYP, cytochrome P450; hERG, human ether-a-go-go-related potassium channel protein; ADMET, absorption, distribution, metabolism, excretion, and toxicity; SAR, structure−activity relationship; ClogP, calculated logarithm of the partition coefficient; log *D*, logarithm of the distribution coefficient; PPB, plasma protein binding; PCA, principal component analysis

REFERENCES

(1) Kennedy, T. Managing the drug discovery/development interface. *Drug Discovery Today* **1997**, *2*, 436−444.
(2) Kubinyi, H. Drug research: myths, hype and reality. *Nat. Rev. Drug Discovery* **2003**, *2*, 665−668.
(3) Schuster, D.; Laggner, C.; Langer, T. Why drugs fail—a study on side effects in new chemical entities. *Curr. Pharm. Des.* **2005**, *11*, 3545−3559.
(4) Kola, I.; Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discovery* **2004**, *3*, 711−716.
(5) DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G. The price of innovation: new estimates of drug development costs. *J. Health Econ.* **2003**, *22*, 151−185.
(6) 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.
(7) Gleeson, M. P. Generation of a set of simple, interpretable ADMET rules of thumb. *J. Med. Chem.* **2008**, *51*, 817−834.
(8) 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.
(9) Waring, M. J. Defining optimum lipophilicity and molecular weight ranges for drug candidates—molecular weight dependent lower log D limits based on permeability. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 2844−2851.
(10) 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.
(11) Hughes, J. D.; Blagg, J.; Price, D. A.; Bailey, S.; DeCrescenzo, G. A.; Devraj, R. V.; Ellsworth, E.; Fobian, Y. M.; Gibbs, M. E.; Gilles, R. W.; Greene, N.; Huang, E.; Krieger-Burke, T.; Loesel, J.; Wager, T.; Whiteley, L.; Zhang, Y. Physiochemical drug properties associated with in vivo toxicological outcomes. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4872−4875.
(12) Waring, M. J.; Johnstone, C. A quantitative assessment of hERG liability as a function of lipophilicity. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1759−1764.
(13) Hill, A. P.; Young, R. J. Getting physical in drug discovery: a contemparary perspective on solubility and hydrophobicity. *Drug Discovery Today* **2010**, *15*, 648−655.
(14) 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 Discovery Today* **2009**, *14*, 1011−1020.
(15) Ritchie, T. J.; Macdonald, S. J. F.; Young, R. J.; Pickett, S. D. The impact of aromatic ring count on compound developability: further insights by examining carbo- and hetero-aromatic and -aliphatic ring types. *Drug Discovery Today* **2011**, *16*, 164−171.
(16) Lovering, F.; Bikker, J.; Humblet, C. Escape from flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **2009**, *52*, 6752−6756.
(17) Yan, A.; Gasteiger, J. Prediction of aqueous solubility of organic compounds by topological descriptors. *QSAR Comb. Sci.* **2003**, *22*, 821−829.
(18) Leeson, P. D.; St-Gallay, S. A.; Wenlock, M. C. Impact of ion class and time on oral drug molecular properties. *Med. Chem. Commun.* **2011**, *2*, 91−105.
(19) Yang, Y.; Chen, H.; Nilsson, I.; Muresan, S.; Engkvist, O. Investigation of the relationship between topology and selectivity for druglike molecules. *J. Med. Chem.* **2010**, *53*, 7709−7714.
(20) Ishikawa, M.; Hashimoto, J. Improvement in aqueous solubility in small molecule drug discovery programs by disruption of molecular planarity and symmetry. *J. Med. Chem.* **2011**, *54*, 1539−1554.
(21) Lu, J. J.; Crimin, K.; Goodwin, J. T.; Crivori, P.; Orrenius, C.; Xing, L.; Tandler, P. J.; Vidmar, T. J.; Amore, B. M.; Wilson, A. G. E.; Stouten, P. F. W.; Burton, P. S. Influence of molecular flexibility and polar surface area metrics on oral bioavailability in the rat. *J. Med. Chem.* **2004**, *47*, 6104−6107.
(22) Prentis, R. A.; Lis, Y.; Walker, S. R. Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964−1985). *Br. J. Clin. Pharmacol.* **1988**, *25*, 387−396.
(23) Hansen, N. T.; Kouskoumvekaki, I.; Jorgensen, F. S.; Brunak, S.; Jonsdottir, S. O. Prediction of pH-dependent aqueous solubility of druglike molecules. *J. Chem. Inf. Model.* **2006**, *46*, 2601−2609.
(24) Delaney, J. S. ESOL: estimating aqueous solubility directly from molecular structure. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1000−1005.
(25) Ran, Y.; Yalkowsky, S. H. Prediction of drug solubility by the general solubility equation (GSE). *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 354−357.
(26) Ran, Y.; He, Y.; Yang, G.; Johnson, J. L. H.; Yalkowsky, S. H. Estimation of aqueous solubility of organic compounds by using the general solubility equation. *Chemosphere* **2002**, *48*, 487−509.29.
(27) Sanghvi, T.; Jain, N.; Yang, G.; Yalkowsky, S. H. Estimation of aqueous solubility by the general solubility equation (GSE) the easy way. *QSAR Comb. Sci.* **2003**, *22*, 258−262.
(28) Johnson, S. R.; Chen, X. Q; Murphy, D.; Gudmundsson, O. A computational model for the prediction of aqueous solubility that includes crystal packing, intrinsic solubility, and ionization effects. *Mol. Pharmaceutics* **2007**, *4*, 513−523.
(29) 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. Model.* **2005**, *45*, 581−590.
(30) A binning scheme with the same number of molecules in each bin was also investigated. The relationships between f_{MF} and Fsp^3 and the aqueous solubility still hold (Figure S12a,b in the Supporting Information).
(31) Yee, S. In vitro permeability across Caco-2 cells (colonic) can predict in vivo (small intestinal) absorption in man—fact or myth. *Pharm. Res.* **1997**, *14*, 763−766.
(32) Camenisch, G.; Alsenz, J.; van de Waterbeemd, H.; Folkers, G. Estimation of permeability by passive diffusion through Caco-2 cell monolayers using the drugs' lipophilicity and molecular weight. *Eur. J. Pharm. Sci.* **1998**, *6*, 313−319.
(33) van de Waterbeemd, H.; Camenisch, G.; Folkers, G.; Raevsky, O. A. Estimation of Caco-2 cell permeability using calculated molecular descriptors. *Quant. Struct.-Act. Relat.* **1996**, *15*, 480−490.
(34) Bergström, C. A. S.; Strafford, M.; Lazorova, L.; Avdeef, A.; Luthman, K.; Artursson, P. Absorption classification of oral drugs based on molecular surface properties. *J. Med. Chem.* **2003**, *46*, 558−570.
(35) Liu, X.; Chen, C.; Hop, C. E. Do we need to optimize plasma protein and tissue binding in drug discovery? *Curr. Top. Med. Chem.* **2011**, *11*, 450−466.
(36) Smith, D. A.; Di, L.; Kerns, E. H. The effect of plasma protein binding on in vivo efficacy: misconceptions in drug discovery. *Nat. Rev. Drug Discovery* **2010**, *9*, 929−939.
(37) Rodgers, S. L.; Davis, A. M.; van de Waterbeemd, H. Time-series QSAR analysis of human plasma protein binding data. *QSAR Comb. Sci.* **2007**, *26*, 511−521.
(38) Valko, K.; Nunhuck, S.; Bevan, C.; Abraham, M. H.; Reynolds, D. P. Fast gradient HPLC method to determine compounds binding to human serum albumin. Relationships with octanol/water and immobilized artificial membrane lipophilicity. *J. Pharm. Sci.* **2003**, *92*, 2236−2248.
(39) Colmenarejo, G.; Alvarez-Pedraglio, A.; Lavandera, J.-L. Cheminformatic models to predict binding affinities to Human serum albumin. *J. Med. Chem.* **2001**, *44*, 4370−4378.
(40) Lobell, M.; Sivarajah, V. In silico prediction of aqueous solubility, human plasma protein binding and volume of distribution of compounds from calculated pKa and AlogP98 values. *Mol. Diversity* **2003**, *7*, 69−87.
(41) Yamazaki, K.; Kanaoka, M. Computational prediction of the plasma protein-binding percent of diverse pharmaceutical compounds. *J. Pharm. Sci.* **2004**, *93*, 1480−1494.
(42) Saiakhov, R.; Stefan, L.; Klopman, G. Multiple computer-automated structure evaluation model of the plasma protein binding

affinity of diverse drugs. *Perspect. Drug Discovery Des.* **2000**, *19*, 133−155.
(43) Viskin, S. Long QT syndromes and torsade de pointes. *Lancet* **1999**, *354*, 1625−1633.
(44) Aptula, A. O.; Cronin, M. T. D. Prediction of hERG K+ blocking potency: application of structural knowledge. *SAR QSAR Environ. Res.* **2004**, *15*, 399−411.
(45) Ekins, S.; Crumb, W. J.; Sarazan, R. D.; Wikel, J. H.; Wrighton, S. A. Three-dimensional quantitative structure-activity relationship for inhibition of human ether-a-go-go-related gene potassium channel. *J. Pharmacol. Exp. Ther.* **2002**, *301*, 427−434.
(46) Keserü, G. M. Prediction of hERG potassium channel affinity by traditional and hologram qSAR methods. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2773−2775.
(47) Shimada, T.; Yamazaki, H.; Mimura, M.; Inui, Y.; Guengerich, F. P. Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. *J. Pharmacol. Exp. Ther.* **1994**, *270*, 414−423.
(48) Guengerich, F. P. CYTOCHROME P-450 3A4: regulation and role in drug metabolism. *Annu. Rev. Pharmacol. Toxicol.* **1999**, *39*, 1−17.
(49) Bruneau, P. Search for predictive generic model of aqueous solubility using bayesian neural nets. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1605−1616.
(50) Neuhoff, S.; Ungell, A.-L.; Zamora, I.; Artursson, P. pH-dependent bidirectional transport of weakly basic drugs across Caco-2 monolayers: implications for drug-drug interactions. *Pharm. Res.* **2003**, *20*, 1141−1148.
(51) Fessey, R. E.; Austin, R. P.; Barton, P.; Davis, A. M.; Wenlock, M. C. The role of plasma protein binding in drug discovery. In *Pharmacokinetic Profiling in Drug Research*, 1st ed.; Testa, B., Krämer, S. D., Wunderlie-Allenspach, H., Folkers, G., Eds.; Verlag Helvetica Chimica Acta: Zurich, Switzerland, 2006; pp 119−141.
(52) Schroeder, K.; Neagle, B.; Trezise, D. J.; Worley, J. IonWorks™ HT: a new high-throughput electrophysiology measurement platform. *J. Biomol. Screening* **2003**, *8*, 50−64.
(53) Gleeson, M.; Davis, A.; Chohan, K.; Paine, S.; Boyer, S.; Gavaghan, C.; Arnby, C.; Kankkonen, C.; Albertson, N. Generation of in-silico cytochrome P450 1A2, 2C9, 2C19, 2D6, and 3A4 inhibition QSAR models. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 559−573.
(54) BioByte ClogP, version 4.3, Claremont, CA.
(55) Pipeline Pilot, version 7.5, Accelrys, San Diego, CA.
(56) Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bull.* **1945**, *1*, 80−83.
(57) JMP, version 7, SAS Institute Inc., Cary, NC.