Generation of a Set of Simple, Interpretable ADMET Rules of Thumb

M. Paul Gleeson*

Computational and Structural Chemistry, GlaxoSmithKline Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, United Kingdom

Received September 10, 2007

A set of simple, consistent structure–property guides have been determined from an analysis of a number of key ADMET assays run within GSK: solubility, permeability, bioavailability, volume of distribution, plasma protein binding, CNS penetration, brain tissue binding, P-gp efflux, hERG inhibition, and cytochrome P450 1A2/2C9/2C19/2D6/3A4 inhibition. The rules have been formulated using molecular properties that chemists intuitively know how to alter in a molecule, namely, molecular weight, logP, and ionization state. The rules supplement the more predictive black-box models available to us by clearly illustrating the key underlying trends, which are in line with reports in the literature. It is clear from the analyses reported herein that almost all ADMET parameters deteriorate with either increasing molecular weight, logP, or both, with ionization state playing either a beneficial or detrimental affect depending on the parameter in question. This study re-emphasizes the need to focus on a lower molecular weight and logP area of physicochemical property space to obtain improved ADMET parameters.

1. Introduction

Drug discovery has become increasingly difficult in the last few decades as a result of the extra development hurdles placed in the path of pharmaceutical research programs, and the perceived reduction in the number of tractable targets. ADMET[a] considerations are a particular concern with issues such as high plasma protein binding requiring extra studies for FDA approval. This is because the free fraction and, hence, the safety margin can change significantly as a result of age or certain disease states[1,2] so must be investigated further. Similarly, further development work is needed should a molecule be shown to be either a hERG[3] or P450 inhibitor[4] due to the possibility of cardiac arrhythmia or drug-drug interactions, respectively.

The pharmaceutical industry has reacted to the changes in the development process by profiling the molecules of interest earlier and more extensively using in vivo and in vitro methods. However, these methods still require a molecule to be synthesized and screened, which has prompted extensive research into in-silico methods that can be used to virtually assess molecules from their molecular structure alone. In-silico models have been developed on almost all the key ADMET assays employed within the pharmaceuticals industry and are reviewed in detail elsewhere.[5–10]

Many of the in-silico ADMET quantitative structure–property relationship (QSPR) models reported in the literature rely on more advanced statistical methodologies such as neural networks, genetic algorithms, decision trees, and support vector machines or complex descriptors such as e-state,[11] BCUT,[12,13] or custom pharmacophoric[14,15] descriptors. While leading to more predictive algorithms in general, mechanistic understanding is limited due to the complicated nature of the models.

To counter the general reduction in interpretability of QSPR models, an attempt was made to demonstrate a set of simple rules of thumb based on large data sets a range of ADMET assays run within GSK (Table 1). To allow for greater understanding, a simple statistical technique was used in combination with three simple properties familiar to medicinal chemists.

Simple rules can often have a greater impact than more complex, more predictive in-silico models due to their greater uptake among the medicinal chemistry community. Examples of these include Lipinski's rule of 5[16] for absorption or guides by van de Waterbeemd et al.[17] for CNS penetration. The former, in particular, has proved immensely useful in lead generation and optimization due to the uncomplicated message given and ease with which a chemist can incorporate the required changes into the next molecule in the series.

2. Results

A small number of information-rich, orthogonal molecular descriptors were required to generate structure–activity relationships (SAR) for the 15 different ADMET assays studied here (Table 1). This subset was obtained by building a principal components analysis[18–22] (PCA) model on ~30000 diverse GSK molecules for which 12 commonly used descriptors had been computed (Figure 1 and Table 2).

Analysis of the PCA model results shows that we can describe 81% of the variation in the 12 descriptors using just four orthogonal components. Components one and two cumulatively describe ~60% of the total variation in the data set, with components three and four describing an additional 21%. Rather than using the four uninterpretable pseudodescriptors, we can alternatively use the descriptor with the largest absolute contribution to that particular component as being representative of the variance explained. Thus, the plot of molecular weight (MWT), the descriptor with the largest weight on component

* To whom correspondence should be addressed. Phone: +44 (0)1438 768682. Fax: +44 (0) 1438 763352. E-mail: paul.x.gleeson@gsk.com.

[a] Abbreviations: ADMET, adsorption, distribution, metabolism, excretion and toxicity; ACD, Advanced Chemistry Development, Inc.; AM, artificial membrane permeability; ANOVA, analysis of variance; clogP, calculated octanol–water logP; CNS, central nervous system; COMFA, comparative molecular field analysis; FDA, Food and Drug Administration; GSK, GlaxoSmithKline; HBA, hydrogen bond acceptors; HBD, hydrogen bond donors; hERG, human ether-a-go-go related gene; HSA, human serum albumin; LG, lead generation; LO, lead optimization; logBB, log of the brain/blood ratio; MDCK, Madine Darby-canine kidney cell; MWT, molecular weight; P450, cytochrome P450; PAMPA, parallel artificial membrane permeability assay; PCA, principal components analysis; P-gp, P-glyco-protein; PSA, polar surface area; QSAR, quantitative structure–activity relationship; QSPR, quantitative structure–property relationship; SAR, structure–activity relationship; USAN, United States assigned name.

10.1021/jm701122q CCC: $40.75
Published on Web 01/31/2008

Table 1. List of the ADMET Assays Studied Herein[a]

category	ADMET assay type	comments	unit
absorption	solubility	phosphate buffer pH 7.4	log(Sol) (mM/L)
	permeability	permeability assay pH 7.4	$\log P_{app}$ (nm/s)
	oral bioavailability	rat - in vivo PK studies	log(*F*)
distribution	volume of distribution	rat - in vivo PK studies	$\log(VD_{ss})$ (L/kg)
	plasma protein binding	rat plasma - equilibrium dialysis	log*K* (%bound/%free)
	CNS penetration	rat - in vivo PK studies	log([brain]/[blood])
	P-gp efflux	MDCK assay with and without a P-gp inhibitor	$\log P_{app}$ (B–A/A–B)
	brain tissue binding	rat brain homogenate - equilibrium dialysis	log*K* (%bound/%free)
metabolism	in vivo clearance	rat - in vivo PK studies	log(CL) (ml/min/kg)
excretion	-	-	-
toxicity	hERG inhibition	[3H] dofetilide assay	pIC_{50}
	1A2 inhibition	fluorescence based assay	pIC_{50}
	2C9 inhibition	fluorescence based assay	pIC_{50}
	2C19 inhibition	fluorescence based assay	pIC_{50}
	2D6 inhibition	fluorescence based assay	pIC_{50}
	3A4 inhibition	fluorescence based assay	pIC_{50}

[a] For further details on the GSK assays employed, see the Experimental Section.

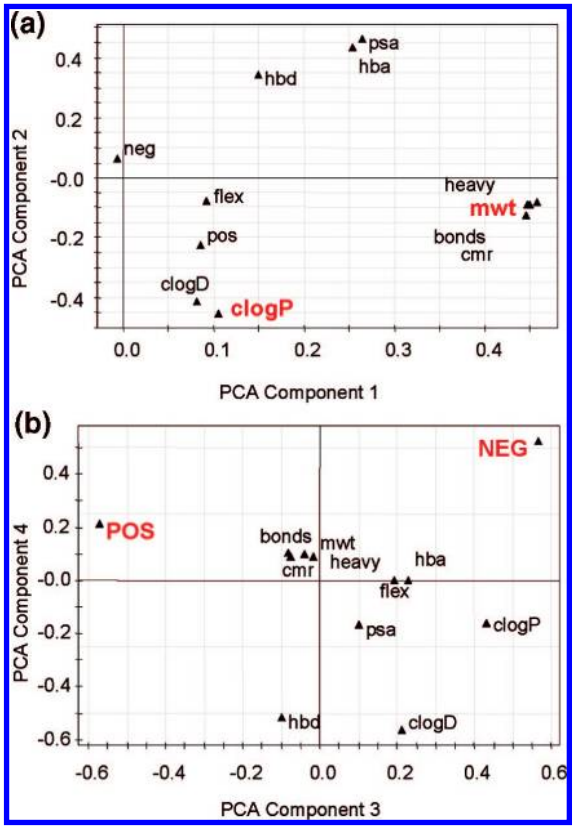

Figure 1. (a,b) PCA loadings plots showing component 1 vs component 2 (top) and component 3 vs component 4. Descriptors at the extreme end of the axes of the loading plots have the greatest impact on the corresponding component. The scores plot can be found in the Supporting Information, Figure S2.

one, against the component in itself, reveals an r^2 of 0.92. This means MWT explains 92% of the total information within component one. For component two, either clogP or the oppositely correlated polar surface area (PSA) could be used. The former was chosen in preference to the latter as it is a more widely used parameter in drug discovery. Note that the correlation between clogP and MWT is just 0.14 for the ~30000 diverse data set, confirming that the parameters are weakly correlated and, thus, essentially orthogonal. Finally, acidic and basic functional group indicator variables are representative of components three and four, respectively, and these can be collectively used to describe the overall ionization state of a molecule (i.e neutral, acidic, basic or zwitterionic[23]). See Figure S1 in the Supporting Information for further details on the relationship between pK_a and the ionization state definition.

Table 2. Selection of In-Silico Descriptors Often Used in Profiling Molecules[23]

No.	descriptors
1	calcd ACD log*D* pH 7.4 (log*D*)
2	calcd ACD log*P* (log*P*)
3	hydrogen bond acceptors (HBA)
4	hydrogen bond donors (HBD)
5	negatively ionizable group (NEG)
6	positively ionizable group (POS)
7	molecular flexibility (FLEX)
8	calcd molar refractivity (CMR)
9	molecular weight (MWT)
10	polar surface area (PSA)
11	rotatable bonds (bonds)
12	heavy atom count (heavy)

The three essentially orthogonal molecular descriptors chosen for this study (size, lipophicity, and ionization state) describe the key bulk property characteristics of molecules and provide us with a means to derive relatively simple rules across a number of ADMET assays. This we discuss in the following section.

2.1. Relationship between Structure and ADMET Liabilities. Analysis of variance (ANOVA) is the statistical method used to assess the size and significance of the differences in the means of the different property groups. The ANOVA results for the 15 different assays are presented graphically in terms of (a) MWT, (b) ionization state, and (c) clogP and ionization state combined. In each ANOVA graph, the mean along with the 95% confidence in the mean for each of the variable categories are illustrated. The effect of a variable on a given ADMET parameter can be considered statistically significant at the 99.9% confidence level unless otherwise stated. Furthermore, the effect of clogP and MWT on the ADMET parameter can be considered independent at the 99.9% confidence level unless otherwise stated.[24]

In each plot provided, the width of the errors bars and the difference in the mean values of the different categories are indicative of the strength of the relationship between the parameters. In the few cases a variable has no effect on a particular ADMET parameter, a graph is still reported, as the lack of a relationship can be as important as the presence of one. Care must be taken to avoid overinterpreting the differences in the means found in these uni and bivariate analyses. For example, we may see a statistically significant difference between the mean ADMET parameters of two different ionization states, but this could in principle be due to, or accentuated

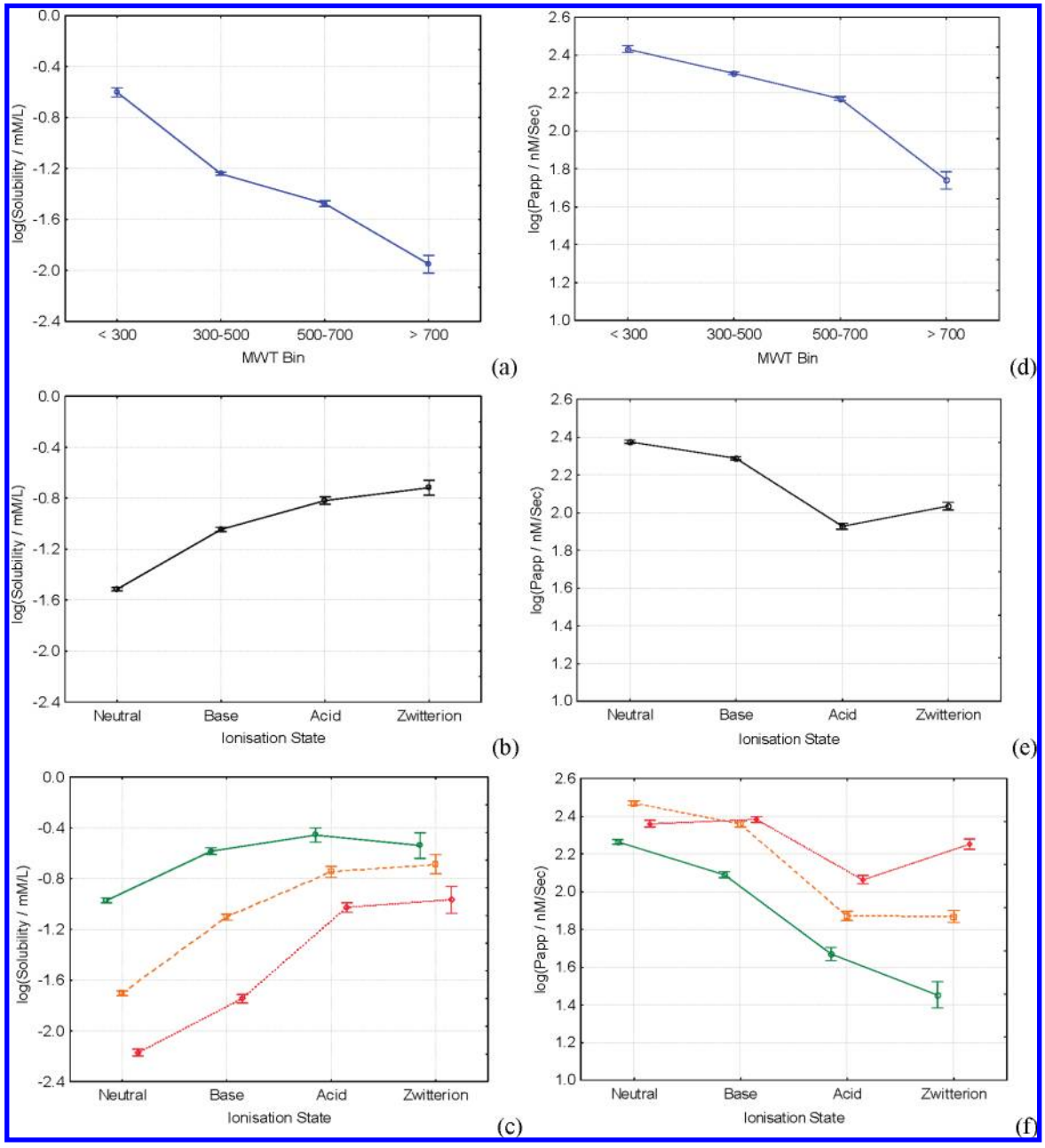

Figure 2. (a−f) Relationship between solubility (a−c) and AM permeability (a−d) with respect to MWT (top), ionization state (middle), and a combination of clogP and ionization state (bottom) for 44584 and 50641 molecules measured in the respective assays. In c and f, green solid line denotes clogPs < 3, orange dashed line denotes clogPs = 3−5, and red dotted line denotes clogPs > 5. The error bars denote the 95% confidence limits of the mean.

by, differences in another property, such as logP. With this in mind, a conservative interpretation of the data is reported, relying on (1) results that pass statistical tests at the 99.9% confidence level, (2) validation of results based on reports in the literature, (3) comparison of the results to those from multivariate in-house analyses,[23] (4) relying on in-house knowledge of the ADMET assays in question, and (5) analysis of the chemotypes found in bins containing few observations or those that appear to be outliers.

2.2. Absorption. 2.2.1. Solubility. Good solubility is an important component of an orally administered drug, determining the amount freely available to permeate through the gastrointestinal membranes into systemic circulation.[25] Considerable emphasis is placed on maximizing this property in lead generation (LG) and lead optimization (LO), relying on both physicochemical assays and in-silico models. To supplement these methods, a set of simple rules of thumb for solubility were determined using phosphate buffer measurements on 44584 diverse molecules (Figure 2).

As the size of a molecule increases, represented here by the MWT, solubility on average decreases in line with reports from the literature[16,26,28,29] (Figure 2a). On average, molecules with MWTs < 300 have solubilities of ∼250 μM (μmol/L), but molecules at the other extreme, with MWTs > 700, have solubilities of just 11 μmol/L. The width of the errors bars, which are dependent on the number of observations in a bin and their standard deviation (σ), indicate that MWT has a large effect on solubility. It should be noted that these results are average effects only and they do not imply that one cannot obtain soluble molecules with high MWTs. The results should be interpreted as being indicative of the likelihood that an optimal value will be obtained given a set of property values.

Ionization state plays a key role in defining the solubility of molecules [28,27] (Figure 2b). The width of the error bars and the range in the solubility averages for the different ionization categories also indicate that ionization has a strong effect. Ionization state would be expected to play an important role on the solubility because the presence of ionizable groups will generally lead to a greater solvation energy increase going from the crystal form to aqueous solution. Zwitterionic molecules contain both an acidic and basic functional group and so are the most highly soluble on average, while neutral molecules are the least soluble on average. Acidic molecules are more soluble on average than basic molecules, as defined here. However, this is more a reflection of the ionization state definition.[23] Molecules defined as acidic are almost exclusively strong acids that are almost fully ionized at pH 7.4. Bases are much more common but show greater variation in their pK_as, even though they are predominantly ionized (>50%) at pH 7.4.

The strong relationship between solubility and lipophilicity is often discussed in the literature, and unsurprisingly, on this large, diverse data set, we find that as clogP increases, solubility on average decreases[16,28–30] (Figure 2c). It is also clear that while neutral molecules are on average more poorly soluble than any other ionization state, lowering the clogP to <3 sees their solubility approach the mean solubility of the ionizable species (Figure 2c). The effect of increasing lipophilicity does not appear to have the same impact on the different ionization states, with the solubility increasing more for neutral and basic molecules compared to acids and zwitterions.

Given that there is often a degree of correlation between lipophilicity and MWT within congeneric series in particular, it is important to demonstrate that the effect they have on solubility is essentially independent. For this data set, the correlation between the two parameters is minimal ($r^2 = 0.096$, $N = 44584$, $P < 0.00001$). Furthermore, an analysis of the effect of clogP on the mean log(Sol) values, for a given MWT bin, indicate that one can be more than 99.9% confident that MWT and clogP have an independent effect on solubility.[24]

2.2.2. Permeability. When an orally administered drug dissolves in the gastro-intestinal tract, it must then be sufficiently permeable through the biological membranes present to enter the systemic circulation. Permeation can occur via transcellular diffusion, paracellular diffusion, and transporter-mediated mechanisms, with the former often being mimicked in the laboratory using artificial membrane assays, such as PAMPA variant,[31] and the latter using MDCK[32] or CACO2[33] cell lines. The principal advantage of the former methodology is that it is less expensive to run, allows greater numbers of molecules to be characterized, and describes the predominant mechanism of permeation.[25] A data set of 50641 unique measurements from an artificial membrane permeability assay (AM) run within GSK were available for analysis.

As the MWT of a particular chemotype increases, membrane permeability on average decreases (Figure 2d). In line with literature reports, it is clear that as a molecule increases in size, its ability to permeate through the artificial mimic of a biological membrane on average decreases.[16,34–36]

Acidic molecules are generally found to be the least permeable through the negatively-charged lipid membranes, and this can be rationalized on electrostatic grounds[9,34] (Figure 2e). Zwitterionic molecules are marginally more permeable than acids on average, essentially a result of the negative charge neutralization that occurs due to the addition of a basic group. Bases are the second most permeable group on average but less than neutral molecules, presumably as a result of the stronger interaction they can form with the anionic membrane itself, thus limiting the rate of diffusion.

Only neutral molecules show the traditional nonlinear dependence between the $\log P_{app}$ and clogP.[36,37] For basic, acidic, and zwitterionic molecules, increasing the clogP appears to lead to an increasing permeability on average (Figure 2f).

2.2.3. Bioavailability. The oral bioavailability of a molecule is a complex pharmaco-kinetic (PK) parameter that comprises absorption (solubility, permeability) and clearance components.[38,39] It is a widely used parameter in drug discovery as it is relatively easy to obtain in vivo as compared to the direct measurement of the fraction absorbed.

In line with reports from others,[16,40] it is found that MWT plays an important role in defining the extent of bioavailability in rat for the $N = 4431$ diverse set of molecules under investigation here (Figure 3a). Molecules with MWTs < 300 have average bioavailabilities of ~18% compared to ~10% for those with MWTs > 700. The impact of MWT on bioavailability is understandable given that both permeability and solubility are strongly dependent on this parameter. However, while statistically significant, the error bars associated with the bioavailability−MWT relationship are larger than those found for solubility or permeability, indicating it has a weaker effect on this parameter.

Ionization state has only a minor impact on the overall bioavailability of molecules (Figure 3b). The average bioavailabilities for neutral, basic and zwitterionic molecules are quite similar, having values of 15, 13, and 10% on average, respectively. The presence of an acidic group on bioavailability is the most surprising, with these molecules being more bioavailable on average (~18%) than any of the other ionization states.[41] This difference could be considered surprising when one recalls that acids are the least permeable of all the ionization states, although this is balanced somewhat by solubility considerations. Additionally, the clearance term, a key constituent of the bioavailability has not been factored in, and it will be shown in section 2.4.1 that the increased bioavailability of acids can in part be attributed to their relatively low clearance in conjunction with good solubility.

The relationship between bioavailability and clogP is not statistically significant at the 99.9% confidence level. This is also the case when broken down by individual ionization state. However, this could be a result of the relatively simplistic modeling performed here[24] (Figure 3c). Alternatively, the lack of a relationship with lipophilicity may not necessarily be surprising given similar reports in the literature[42] and because bioavailability is dependent on a number of different ADMET processes (solubility, permeability, and clearance), each with a subtly different dependence on lipophilicity. It might be expected that reasonable correlations between logP and bioavailability will be observed for particular chemotypes, where the differences in bioavailability are restricted to differences in solubility for example.

It should be noted that PSA has been shown to be important in describing bioavailability elsewhere.[42,43] This is somewhat understandable given that we can see that this parameter is essentially a hybrid of MWT and clogP based on the principal component plot discussed earlier (Figure 1).

2.3. Distribution. 2.3.1. Volume of Distribution. How widely a drug molecule is distributed throughout the body once it is absorbed must be considered, as this can determine whether it will elicit a pharmacological response or not. The volume of distribution is a nonphysiological term that is a measure of drug distribution, which together with clearance determines the half-

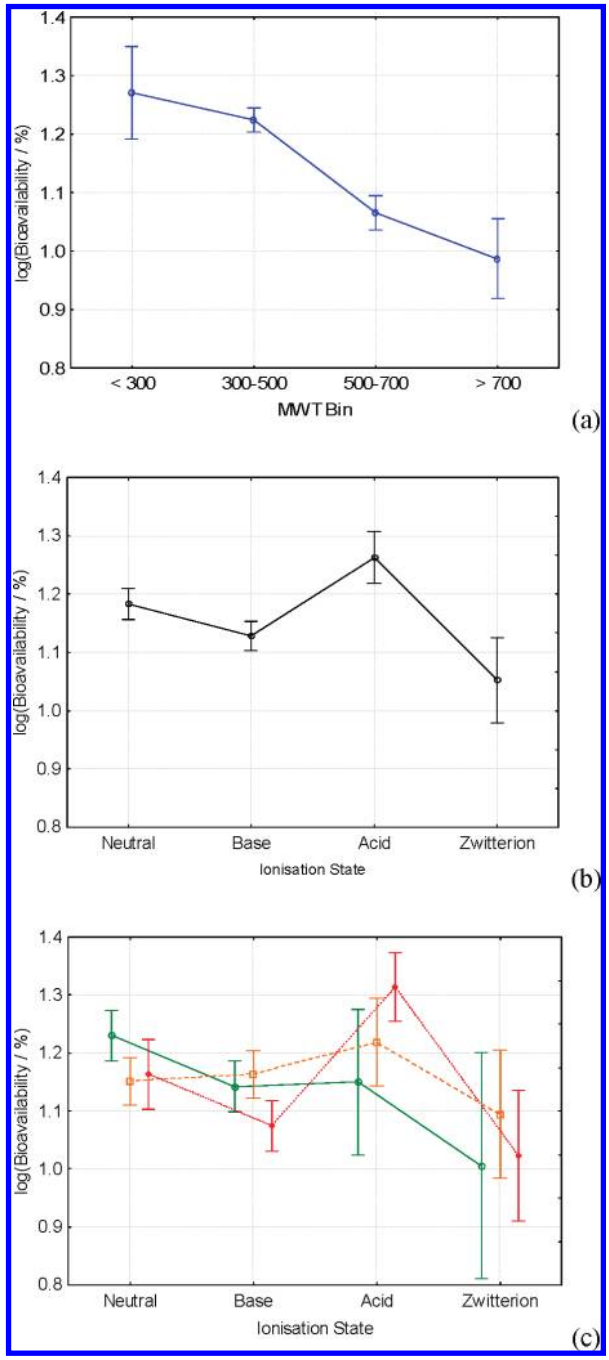

Figure 3. (a−c) Relationship between rat bioavailability and MWT (top), ionization state (middle), and a combination of clogP and ionization state (bottom), for 4431 molecules with measured values. The coloring system is the same as that in Figure 2.

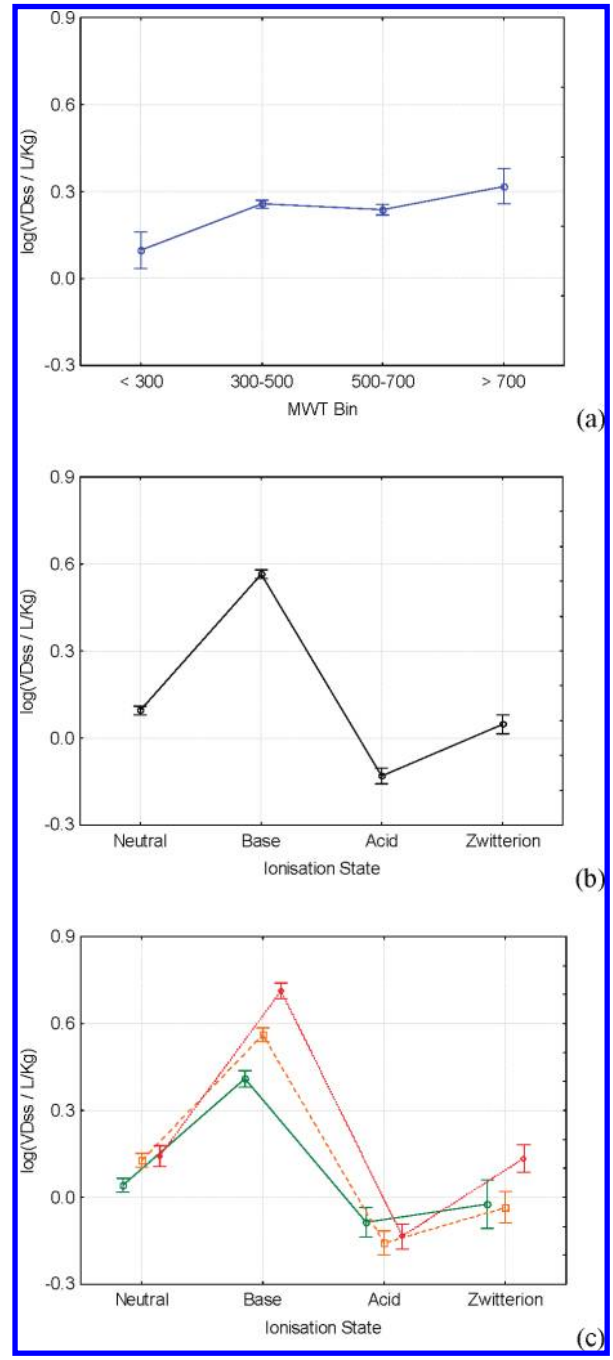

Figure 4. (a−c) Relationship between rat volume of distribution and MWT (top), ionization state (middle), and a combination of clogP and ionization state (bottom) for 9375 molecules with measured values. The coloring system is the same as that in Figure 2.

life and, thus, affects the dosing regimen. Potent, bioavailable compounds may not be effective if they are insufficiently distributed or have too short a half-life, because the concentration of drug maintained at the target might be insufficient to sustain the pharmacological response between doses.

Analysis of the relationship between $\log VD_{ss}$ and MWT for the 9375 observations available shows that increasing the latter will lead to an on average increase in the former (Figure 4a). However, it appears to be a very weak effect based on the small range in the mean values for the different MWT categories and the fact that the effect is limited to the relatively small number of observations in the two extreme MWT bins (MWT > 700 and < 300).

Ionization state has the most dramatic effect on the volume of distribution of molecules, and the trends found are in line with reports in the literature by others[44–46] (Figure 4b). Basic molecules are more widely distributed throughout the body compared to the other ionization states. Neutral and zwitterionic molecules have similar mean $\log VD_{ss}$ values, which are in turn higher than acids. These results can be rationalized on the grounds that acids are generally confined to plasma due to strong human serum albumin (HSA) binding (see Plasma Protein Binding section). Bases do not bind as strongly to plasma proteins and will distribute more widely into other compartments in the body due to their affinity for negatively charged membranes/tissues.

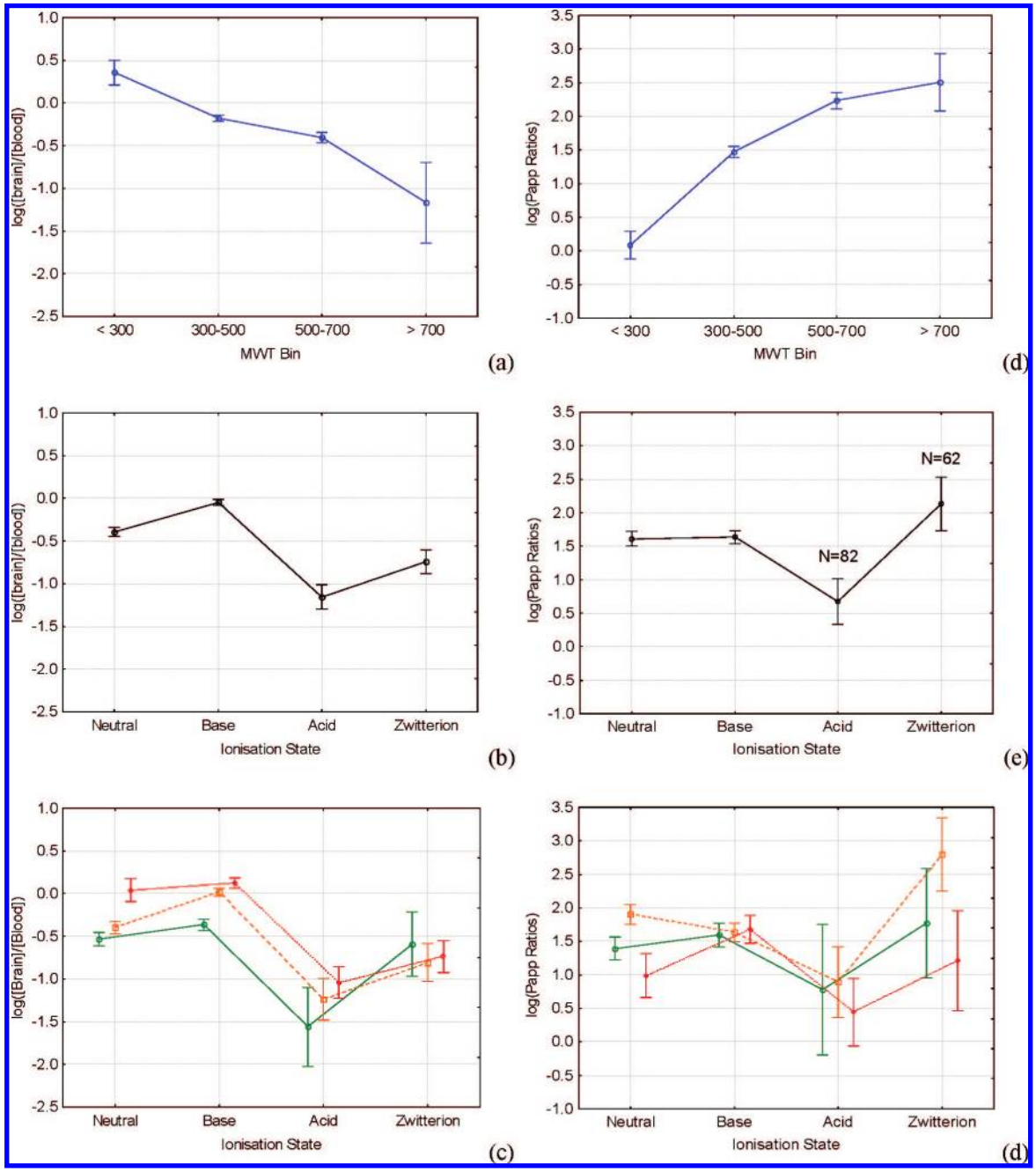

Figure 5. (a−f) Relationship between CNS penetration (a−c) and P-gp efflux ratio (d−f) with respect to MWT (top), ionization state (middle), and the combination of clogP and ionization state (bottom) for 3059 and 1975 molecules measured in the respective assays. The coloring system is the same as that in Figure 2.

It appears that increasing clogP will generally lead to an increase in the volume of distribution of neutral and basic compounds in the body that is in general agreement with reports elsewhere[44,46] (Figure 4c). The effect is not seen for acids and zwitterions, however.

2.3.2. CNS Penetration. For orally administered central nervous system (CNS) drugs to be effective, both oral and brain exposure are required to obtain a pharmacological response. For non-CNS drugs, in contrast, it is generally a requirement that brain exposure is minimized due to the possibility of undesired pharmacological events. To enter the CNS, a molecule must first traverse the blood−brain barrier, with transcellular and active transport mechanisms, such as P-glycoprotein (P-gp) efflux, being more important as a result of the tight junctions between cells.[47]

From an analysis of 3059 diverse molecules with CNS penetration data, we find that as MWT increases CNS penetration on average decreases (Figure 5a). The effect is both large and statistically significant, even taking into account the MWT > 700 bin, which has only 11 observations. Molecules with MWTs < 300 have brain/blood ratios of 2.2 compared to 0.1 for molecules with MWTs > 700. Thus, as a molecule gets bigger, its ability to permeate into the central nervous system decreases, which is in line with reports by Gerebtzoff et al.,[48] who showed similar relationships with molecular cross-sectional area using a much smaller data set. The results are also in line with the findings from the artificial permeability assay analyses reported herein, as would be expected (Figure 2d).

Ionization state plays a key role in defining the extent of CNS penetration of molecules (Figure 5b). Basic molecules are on

average more CNS penetrant than neutral molecules, followed by zwitterions. Acidic molecules are the least CNS penetrant of the four ionization states, in line with the permeability measures described herein and a CNS QSPR study by Platts et al.[49]

The key difference observed between CNS penetration and permeability is that bases are found to be more CNS penetrant than neutral molecules, in contrast to AM permeability. However, this does not take into account differences in other key properties that might cloud the picture, including the pH gradient the AM assay is run at. By taking into account lipophilicity, it is apparent that the difference in penetration between neutral and basic molecules identified in Figure 5b is minimal (Figure 5c). Furthermore, Platt et al. showed that an acidic indicator variable added extra value in predictive equations but a basic term did not,[49] suggesting neutral and basic molecules behave the same with respect to CNS penetration. The result for CNS penetration is still subtly different compared to that for the AM assay, where the neutral and bases are concerned, the former molecules being more permeable on average than the latter for a given clogP. It is difficult to say if this is a real difference or a reflection of the AM assay.

In general, it is found that as the clogP of a molecule increases, the CNS penetration on average increases, which is understandable given that lower PSA values have been found to increase CNS penetration based on other analyses.[17,50] Nonetheless, the effect of MWT is stronger, as can be seen from the range of the mean values for each MWT category when compared to the former. This suggests lower MWT compounds with a correspondingly lower PSA (or higher clogP) are desired for improved CNS penetration. Breaking it down by ionization state reveals a similar trend for neutral, basic, and zwitterionic molecules but not zwitterions, which could be a result of the small numbers of observations present.

Analysis of the effect of clogP on the mean logBB values, beyond any MWT differences, show the two parameters have a statistically significant independent effect on CNS penetration at the 99.1% confidence level, below the threshold 99.9% benchmark value used here. This result is above the more lax but still commonly used 95% confidence level so one can still say with a good degree of confidence that size and lipophilicity have an independent effect on CNS penetration.[24]

2.3.3. P-gp Efflux. P-gp is an important transporter protein found in cells throughout the body, such as those lining the intestine and the blood−brain barrier.[51] P-gp is believed to play an important role in defining the extent of distribution of drug molecules as a result of its ability to remove a structurally diverse range of molecules from compartments in the body.[52] P-gp efflux is believed to play a role in limiting oral exposure and brain exposure of compounds, in particular, so is of considerable importance to the pharmaceuticals industry.

Reports from the literature suggest that increasing molecular bulk/weight/surface area is important in the defining the P-gp efflux in various data sets,[53–55] and the effect can also be seen from 1975 diverse molecules with measured values in GSK assays. As MWT increases, the P-gp efflux ratio (Papp B–A: A–B) on average increases (Figure 5d). From the analyses of permeability and CNS penetration data, we know that higher MWT leads to lower permeability/penetration on average; therefore, an added problem is that these molecules are more likely to suffer from efflux by P-gp. Thus, as a molecule gets bigger, the permeability on average decreases, limiting oral and/or brain exposure, and P-gp efflux becomes an additional complication reducing exposure.

Ionization state plays only a minor role in determining the extent of efflux when compared to MWT. Neutral and basic molecules have similar mean efflux ratios, with zwitterionic molecules having noticeably higher and acidic molecules having noticeably lower means (Figure 5e). This result is not surprising given that the number of hydrogen bond donors[51] is reported as being important for substrate recognition rather than any particular ionization state.

Lipophilicity has a weak nonlinear effect on the P-gp efflux ratio. Molecules with clogPs between 3 and 5 have higher mean efflux ratios than those with clogPs < 3 or > 5. Increasing lipophilicity has been implicated in a number of studies relating structure to P-gp efflux.[54,51] Broken down by ionization state, one finds that neutral molecules display a nonlinear relationship with clogP, while the efflux ratio for basic molecules increase only marginally, but linearly, with increasing clogP (Figure 5f). The small number of acidic and zwitterionic molecules in the data set make it difficult to obtain reliable trends for these species. However, one can be confident (99.1% level) that MWT and clogP have an independent effect on the P-gp efflux.

2.3.4. Plasma Protein Binding. The binding of drugs to plasma proteins, and to human serum albumin (HSA), in particular, has widespread ADMET implications. Plasma protein binding affects the clearance according to the Well Stirred Model,[56] the volume of distribution according to the Gillette equation,[57] as well as efficacy, because it is the free fraction of the drug that is required to elicit a pharmacological response. As such, considerable effort is placed on the prediction of this parameter in vitro and in-silico within the pharmaceutical industry.

From an analysis of the 2939 diverse molecules with in vitro plasma protein binding data, one can see that as MWT increases plasma protein binding on average increases[58] (Figure 6a). On average, molecules with MWTs < 300 are 72% bound and molecules with a MWT between 300 and 500 are 54% bound, while molecules with MWTs between 500 and 700 are 98.2% bound. Because only 13 observations are found in the ">700" bin, it is not possible to say with any confidence whether there is a drop in binding affinity as MWT increases above 700.

In terms of ionization state, binding to plasma proteins follows the trend acids $>$ neutrals $>$ zwitterions $>$ bases (Figure 6b), which is in line with reports of Davis et al. based on a plasma-derived data set[59] and Valko et al.[60] on HSA binding data set.

Lipophilicity is a key contributor to the extent of binding, the process having a strong nonspecific component (Figure 6c). Thus, as clogP increases, plasma protein binding on average increases, in line with reports from others.[23,60–65] It can also be seen that, while bases and zwitterions are less protein bound than acids, they can achieve equivalent levels of binding should the clogP be sufficiently high.

The multivariate PLS model results reported by the author on the same data set[23] showed that MWT had a significant, independent effect on protein binding above lipophilicity. However, from these simplistic ANOVA models, the effect of clogP is significant at the 96.2% confidence level. This result also highlights that certain relationships might be missed as a result of the simple, albeit interpretable modeling when allied with quite rigorous statistical confidence limits.[24] However, given that this is above the commonly used 95% value and with the knowledge of the multivariate results on the same data set, we can be reasonably confident that the parameters have an independent effect on binding.

2.3.5. Brain Tissue Binding. Brain tissue binding is a key ADMET parameter that has implications for the efficacy of CNS

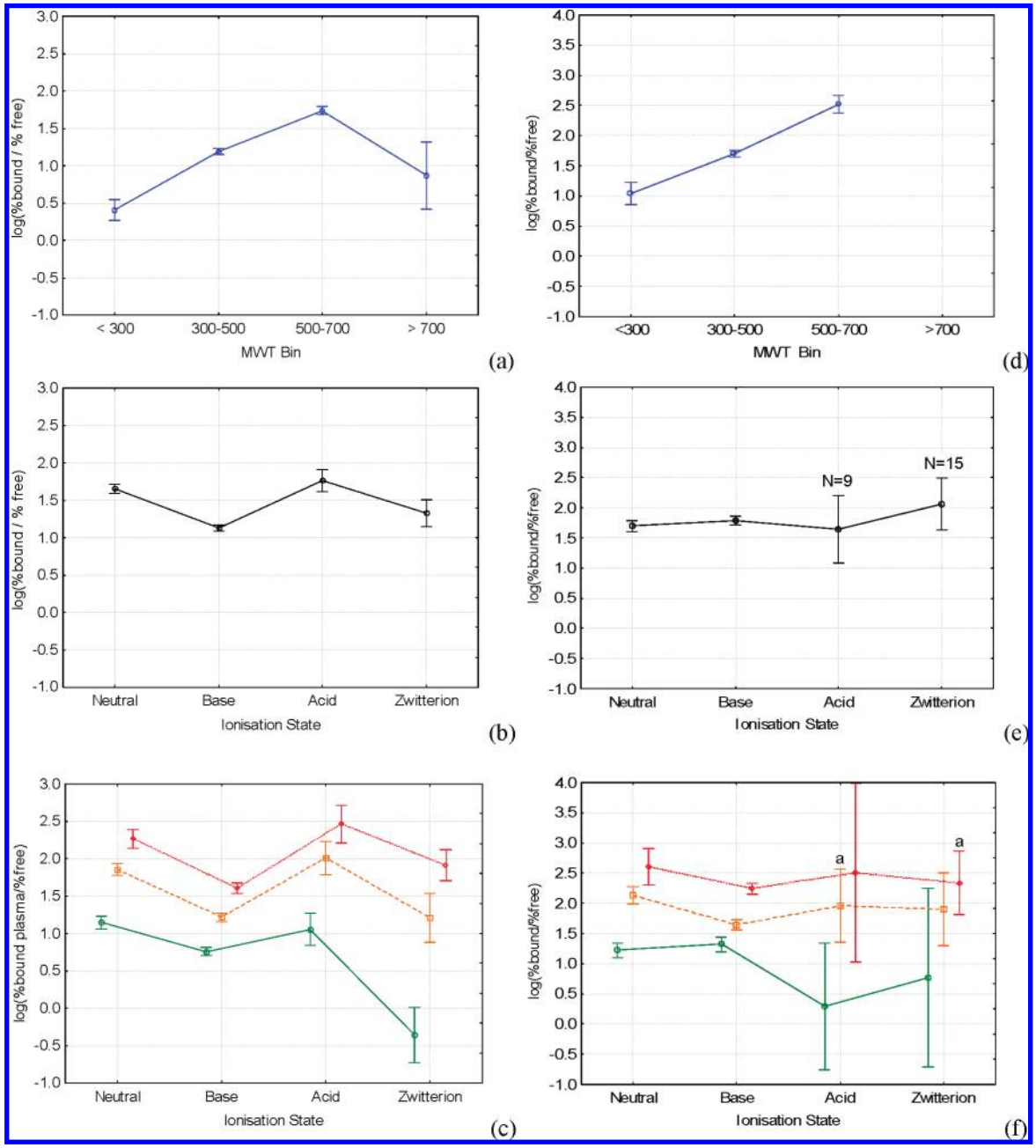

Figure 6. (a−f) Relationship between plasma protein binding (a−c) and brain tissue binding (d−f) with respect to MWT (top), ionization state (middle), and the combination of clogP and ionization state (bottom) for 2939 and 986 molecules measured in the respective assays. The coloring system is the same as that in Figure 2. [a]Although the results reported in (e) and (f) are statistically significant, they are complicated by the small number of observations in some categories. These issues are described in the text.

drugs because, like plasma protein binding, it is the free fraction in the brain that will elicit a pharmacological response.[66] Thus, it is important to consider the effect of molecular properties on both the extent of CNS penetration and the extent of brain tissue binding and whether they show different dependencies.

An analysis of 986 brain tissue binding measurements shows that as MWT increases, brain tissue binding on average increases (Figure 6d). Comparison of the results to those in plasma (Figure 6a) reveals distinct similarities, with larger molecules binding more strongly to both types of tissue. Like the plasma protein binding data set, no lessons can be learned from the MWT > 700 bin, in this case, because no observations are present.

Brain tissue binding does not seem to be dramatically affected by ionization state, as compared to plasma protein binding. Nevertheless, the small differences in the mean for the given ionization states are still statistically significant, at the 99.7% level (Figure 6e). The differences in logK (%bound/%free) for neutral and basic molecules are small compared to that found in plasma, while the small number of acids and zwitterions makes it difficult to draw reliable conclusions. Austin et al.[67] reported that the binding of bases was stronger to hepatocytes compared to neutral or acidic molecules for a given logP, so the differences observed here may simply be a reflection of the tissue differences.

From Figure 6f, it is apparent that binding to brain tissue has a similar dependence on lipophilicity as plasma protein binding. Breaking the data down by ionization state reveals no difference, suggesting brain tissue binding is driven primarily by nonspecific effects. This is in line with brain tissue binding studies reported on a set of basic molecules,[68] as well as

conforming to our understanding of nonspecific binding to other tissues such as plasma proteins and hepatocytes.[67] Additionally, one can be confident (96.1% level) that MWT and clogP have a significant independent effect on brain tissue binding.

2.4. Metabolism. 2.4.1. In-vivo Clearance. Molecules that possess good solubility and permeability may have low oral bioavailability if the in vivo clearance (or plasma/blood clearance) is too high. The in vivo clearance is also important as it determines the half-life of a molecule in conjunction with the volume of distribution, having implications for the dosing interval. Clearance can occur via hepatic, renal, and bilary processes and is considered to be the most difficult of the ADMET process to predict due to the stronger dependence on structural aspects than physicochemical properties.[69]

Analysis of the 11490 diverse molecules with clearance measurements shows that there is no meaningful relationship between MWT and the logCL (Figure 7a). Moving through the MWT bins in increasing order, the mean clearances are 25.7, 25.1, 29.5, and 22.9 mL/min/kg, and while these differences are technically statistically significant above the 99.9% confidence level, they are of little importance from an SAR viewpoint. This finding is understandable given the lack of even simple QSPR models published in the literature, which might be explained by the large number of different bioactivation pathways associated with metabolism.[70]

Ionization state plays a more important role in defining the extent of clearance in vivo. From Figure 7b, one can see that the mean logCL of acidic molecules are considerably lower on average than neutral and zwitterionic molecules, which in turn are lower than bases. The range of the means and the width of the errors bars indicate that this is a significant effect. These differences in clearance can be rationalized based on protein binding considerations, which are known to be an important factor in determining the extent of clearance[71,72] (i.e., the amount of drug free in plasma will affect how highly cleared a molecule is). Acidic molecules are the most highly bound on average and, thus, least susceptible to clearance in the liver, while bases are the least highly bound and are the most highly cleared.

A very weak nonlinear relationship appears to exist between clogP and the in vivo clearance, which is statistically significant. The strength of the relationship can be appreciated by considering the mean values associated with each clogP bin: 24.6 mL/min/kg for clogP < 3, compared to 27.5 for clogP $= 3$–5 and 26.9 for clogP > 5. The dependence on clogP differs subtly when broken down by ionization state (Figure 7c). Increasing the clogP has only a small but generally increasing effect on the clearance of neutral and basic molecules. In contrast, the in vivo clearance of acids and zwitterions decreases marginally with increasing clogP bin.

Collectively, these results suggest that a greater emphasis on structural considerations will be needed to generate reliable SAR for in vivo clearance, at least compared to more physical property controlled ADMET processes.

2.5. Toxicity. 2.5.1. hERG Inhibition. Inhibition of the voltage gated potassium ion channel, a trans-membrane protein encoded by the hERG gene, is known to be undesirable due to the possibility of QT prolongation, which can lead to fatal cardiac arrhythmia.[73] This has led to the development of in vitro assays within the pharmaceuticals industry to screen for the problem early on in the development process.[74]

From an analysis of the 35200 molecules screened in a hERG inhibition assay in GSK, we can see that, as MWT increases, the mean pIC_{50} on average increases (Figure 8a). These findings are in line with QSAR studies by Aptula et al.,[75] who showed

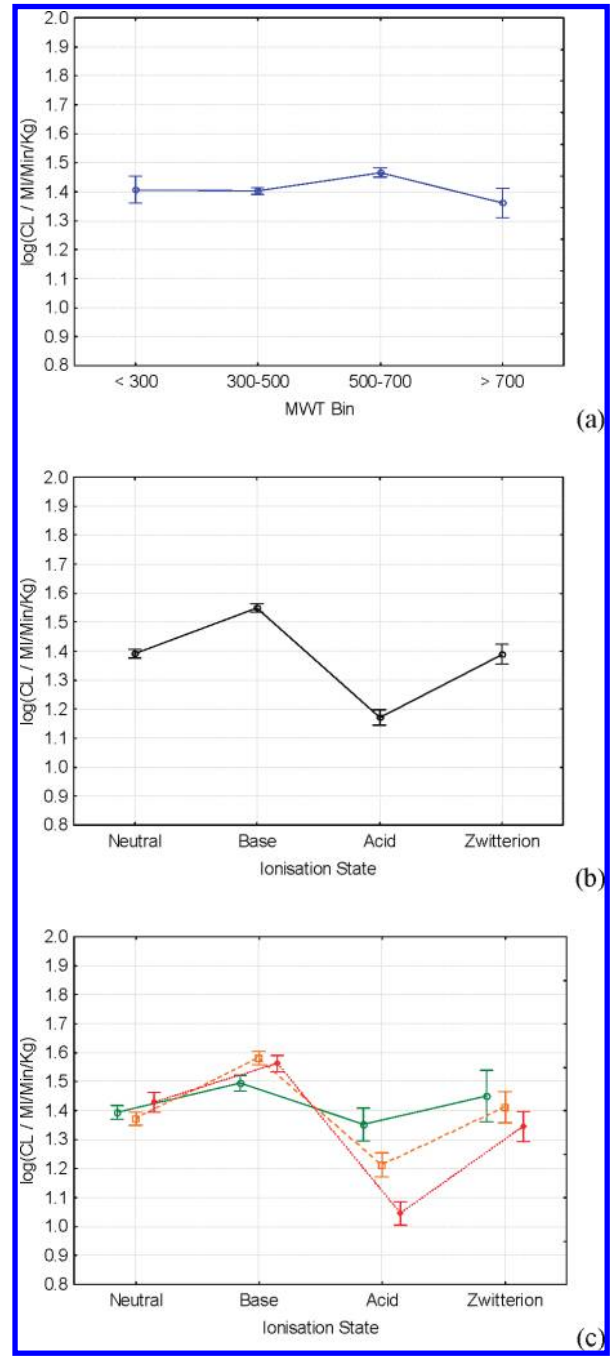

Figure 7. (a−c) Relationship between in vivo clearance and MWT (top), ionization state (middle), and the combination of clogP and ionization state (bottom) for 11490 molecules measured in the respective assays. The coloring system is the same as that in Figure 2.

that the maximum distance in a molecule, a size indicator, in conjunction with logD, gave qualitative discrimination of a set of 19 hERG inhibitors. Furthermore, pharmacophoric models generated by Ekins et al.[76] contained four hydrophobic features and models by Keseru[77] contained three, suggesting larger molecules have a greater tendency to occupy the hERG active site.

Ionization state is also important in defining the types of compounds that inhibit hERG, which might be expected due to its role in transporting K^+ ions. Neutral and acidic[78,79] molecules have lower hERG inhibition on average compared to basic or zwitterionic molecules, which is in agreement with the findings of others that a strong basic center is generally required for potent hERG inhibitors[15,75–80] (Figure 8b).

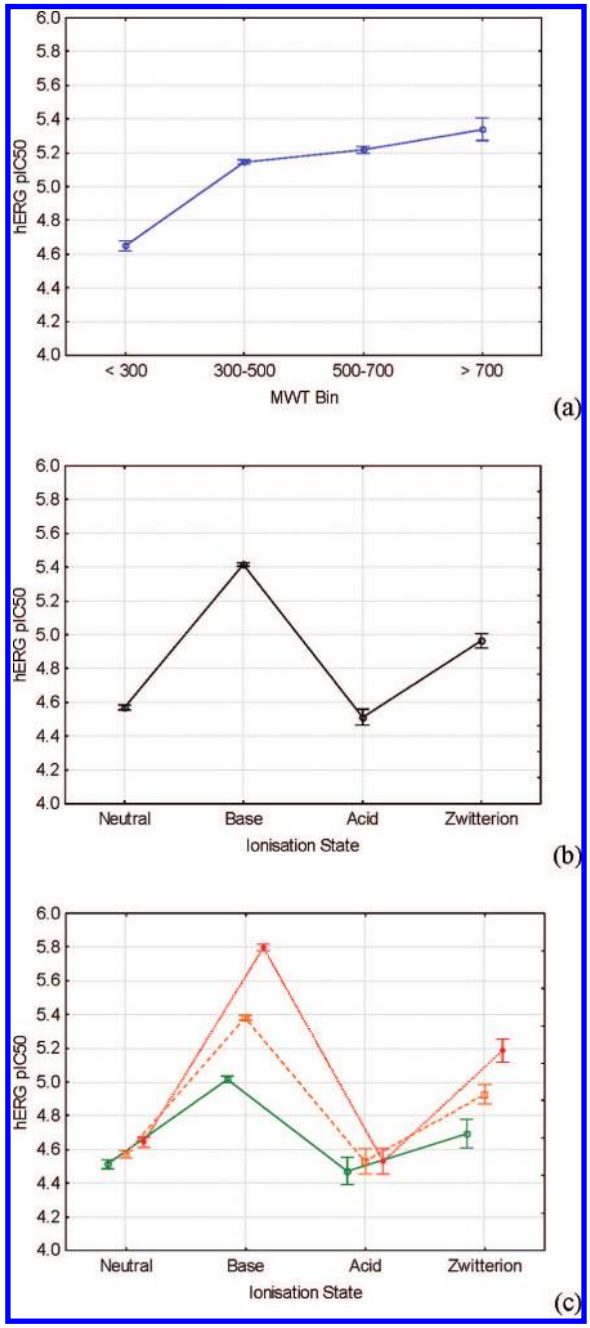

Figure 8. (a−c) Relationship between hERG inhibition and MWT (top), ionization state (middle), and a combination of clogP and ionization state (bottom), for 35200 molecules with measured values. The coloring system is the same as that in Figure 2.

Also in line with a number of literature studies is the finding that as clogP increases, the hERG pIC_{50} on average increases,[75–77] and this difference is not simply due to increases in MWT. Breaking the data down by ionization state reveals distinct differences; the mean pIC_{50}s of neutral and acidic molecules do not change dramatically with increasing clogP bin, in contrast to basic and zwitterionic molecules (Figure 8c). This might be expected, because the basic pharmacophoric feature has not been matched. It is also apparent from Figure 8c that basic molecules will have a greater liability on average, irrespective of the clogP. For example, basic molecules with a clogP > 5 have a mean pIC_{50} of 5.8, compared to 5.4 when the clogP is between 3 and 5 and 5.0 when the clogP < 3. In contrast, the mean pIC_{50} of clogP > 5 neutral molecules is just 4.6.

2.5.2. Cytochrome P450 Inhibition. Cytochromes P450s are a family of diverse heme-containing proteins that catalyze the metabolism of a broad range of molecules of both exogenous and endogenous origin[81–83] and are implicated in toxicological events.[84] Should two or more drugs be administered concomitantly, with one being a potent P450 substrate or inhibitor,[84] the metabolism of the other drug may be inhibited, causing its plasma levels to rise, which may lead to undesirable toxic effects. An additional complication is that a molecule that is shown to be P450 substrate may in fact pose a toxicity risk, not a result of the parent molecule itself, but due to the reactive metabolite formed.

The inhibition of cytochrome P450s 1A2, 2C9, 2C19, 2D6, and 3A4 are the focus here, with these isoforms making up approximately 60% of the total hepatic CYPs[85,86] and being accountable for the metabolism of the majority of known drugs.[87] While the results reported here are for P450 inhibitors, it might be expected that we can generalize the results to substrates or inducers also given that these must typically bind effectively within the active site and, thus, should show common trends.[88]

2.5.2.1. P450 1A2 Inhibition. Cytochrome P450 1A2 accounts for approximately 13% of hepatic P450s[86] and is implicated in the metabolism of drug molecules such as paracetamol and phenacetin, as well as caffeine.[89] From an analysis of the 49837 molecules with measured 1A2 pIC_{50}s one finds that as the MWT increases the pIC_{50} on average decreases (Figure 9a). These results indicate that increasing the size of a molecule will on average lead to a decrease in 1A2 liability, in contrast to the other CYP isoforms considered herein. This finding is in agreement with the recently reported crystal structure of 1A2, which displays a narrow cavity of limited volume compared to other isoforms.[90] The results are also broadly in line with the results of Burton et al.[91] who found that size and aromaticity were the key features discriminating inhibitors and noninhibitors. Collectively, these results suggest that steric factors associated with the 1A2 active size restrict all but the smallest molecules from binding effectively.[91,92]

Ionization state plays a relatively small role in determining the extent of 1A2 inhibition when compared to MWT (Figure 9b) and this can be rationalized by the lack of ionized residues capable of ligand interaction within the active site crystal structure.[90] The differences in the means, while small, are all statistically significant due to the large number of observations in each category. Neutral molecules have the highest 1A2 pIC_{50}s on average, followed by bases, acids, and finally zwitterions.

Descriptors describing the ionization state have not appeared in 1A2 QSAR models, which may help to confirm the findings above. Furthermore, from COMFA models built on 52 naphthalene, lactone, and quinoline derivatives, Korhonen[93] concludes that the 1A2 active is primarily composed of hydrophobic and aromatic residues. This might explain the small variation in the mean pIC_{50}s for the different ionization states found in Figure 9b.

Lipophilicity has been implicated in 1A2 inhibition elsewhere in the literature,[93] however, here we can see that it has a relativity weak effect on the extent of inhibition compared to MWT (Figure 9c). clogP also has a subtly different effect on the different ionization states, with neutral molecules having the largest mean pIC_{50} when the clogP is between 3 and 5, basic molecules having the maximum liability when the clogP is >3 and acids having the maximum liability when the clogP > 5. The situation is less clear for zwitterionic molecules due to the relatively small number of observations and due to confounding

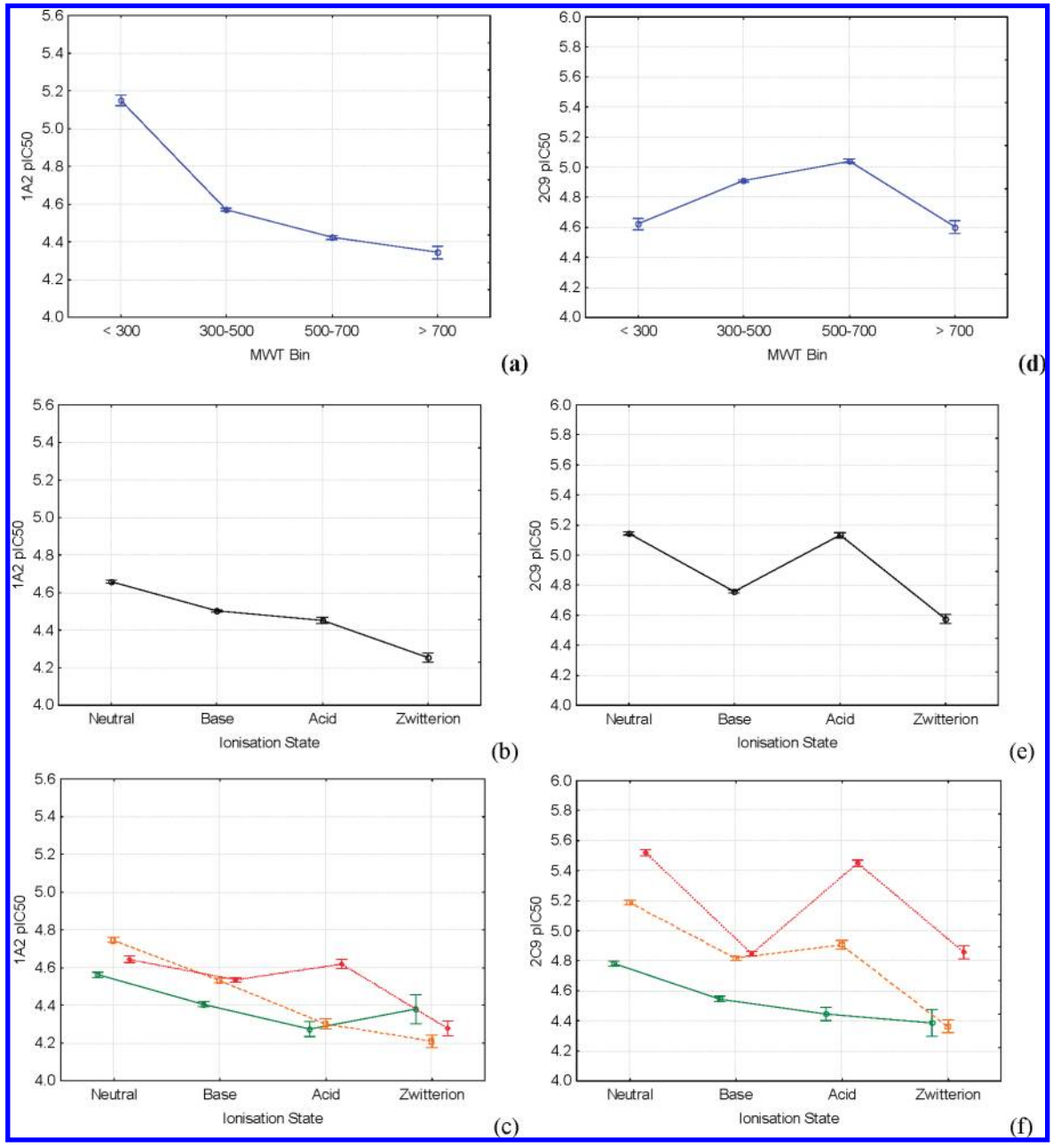

Figure 9. (a−f) Relationship between CYP-1A2 inhibition (a−c) and CYP-2C9 inhibition (d−f) with respect to MWT (top), ionization state (middle), and a combination of clogP and ionization state (bottom) for 49837 and 51097 molecules measured in the respective assays. The coloring system is the same as that in Figure 2.

MWT effects in the zwitterionic clogP < 3 bin as a result of certain chemotypes.

2.5.2.2. P450 2C9 Inhibition. The cytochrome P450 2C family accounts for approximately 20% of hepatic P450s,[86] and cytochrome P450 2C9 in particular is implicated in the metabolism of drug molecules such as phenytoin, tolbutamide, and warfarin.[94] From an analysis of 51097 molecules with measured 2C9 pIC_{50}s one finds that MWT plays an important role in determining the extent of 2C9 inhibition. Unlike hERG, which shows a linear increase in affinity with MWT, or 1A2, which shows a linear decrease, 2C9 shows a parabolic effect. Molecules with MWTs between 300 and 700 are the most potent on average at 2C9, suggesting steric factors associated with the 2C9 active size generally prevent very large or very small molecules from binding optimally (Figure 9d).

Ionization state plays an important role in determining the extent of 2C9 inhibition. Neutral and acidic molecules[94] have the highest 2C9 affinity on average (Figure 9e), followed by bases and zwitterions. From crystallography data, Williams et al.[95] reported that no basic residues critical for binding the acidic warfarin molecule were found in the 2C9 active site to explain the increased affinity of acids. However, Wester et al. report that Arg108 plays an important role in binding the acidic molecule flurbiprofen,[96] this being found in a catalytically active conformation, unlike the former.

In terms of lipophilicity, one finds that as clogP increases 2C9 inhibition on average increases[94] and this difference is not simply due to increases in MWT. Molecules with clogP < 3 have a mean pIC_{50} of 4.6, those with clogP between 3 and 5 have a mean value of 4.9, while those with clogP > 5 have a

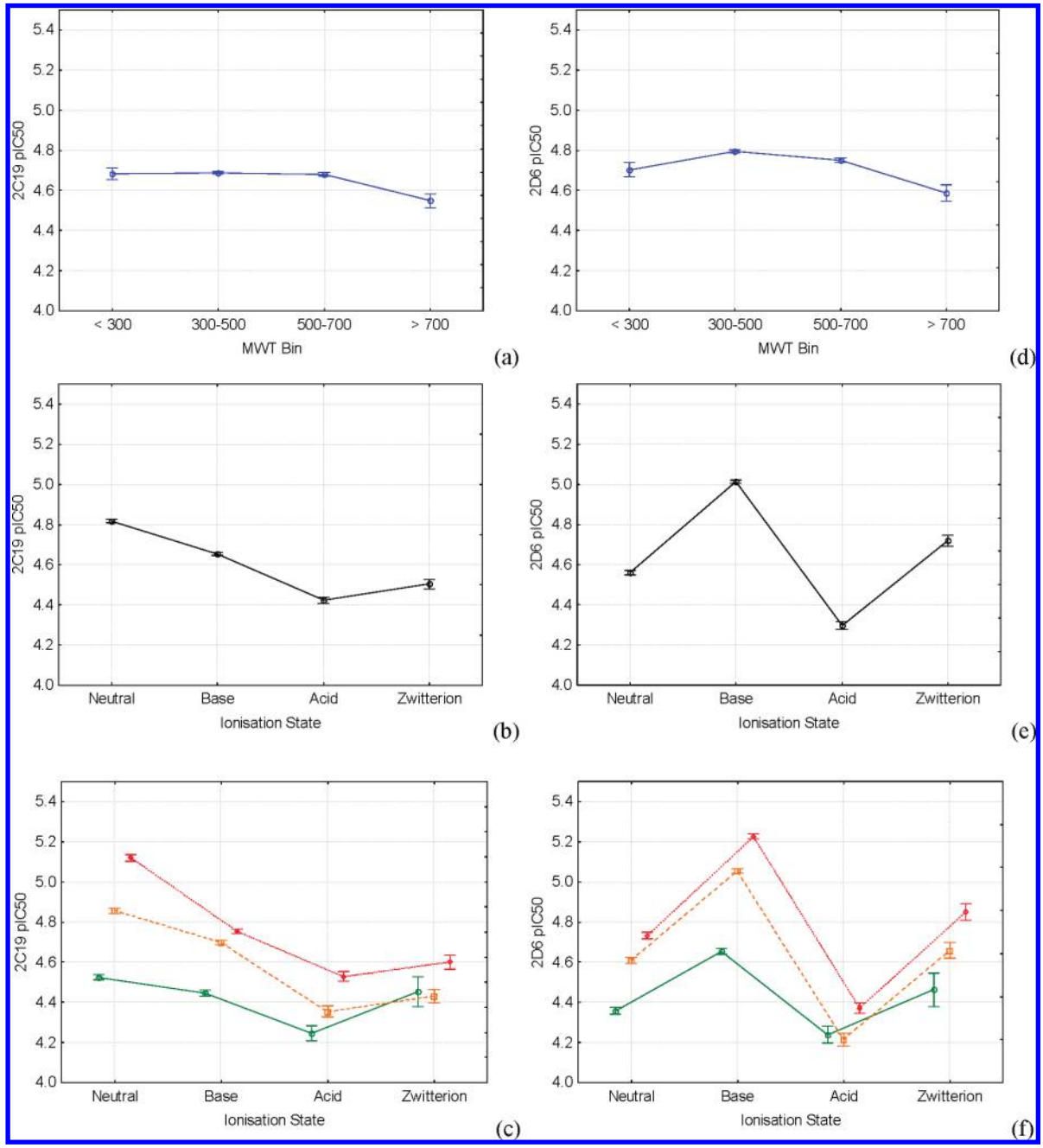

Figure 10. (a−f) Relationship between CYP-2C19 inhibition (a−c) and CYP-2D6 inhibition (d−f) with respect to MWT (top), ionization state (middle), and a combination of clogP and ionization state (bottom) for 48464 and 50886 molecules measured in the respective assays. The coloring system is the same as that in Figure 2.

mean of 5.2. The dependency between 2C9 inhibition and clogP is broadly the same when broken down by ionization state (Figure 9f), however, the magnitude of the potency change is more pronounced for neutral and acidic molecules compared to bases or zwitterions.

2.5.2.3. P450 2C19 Inhibition. Cytochrome P450 2C19 is also a member of the 2C family[86] implicated in the metabolism of drugs molecules such as omeprazole, propranalol, and diazapam.[94] From an analysis of the SAR derived from 48464 molecules with measured 2C19 pIC_{50}s one finds that inhibition shows little dependence on MWT, even though this isoform is structurally very similar (91%)[97] to 2C9 discussed in the previous section (Figure 10a). These results suggest the steric requirements imposed on substrates/inhibitors by the 2C19 active size are very subtle compared to those found for the other isoforms considered here. It may alternatively mean that molecular recognition is more important for determining the affinity for 2C19.

The ionization state of a molecule also seemed to be of only minor importance in determining the extent of inhibition of 2C19. Neutral molecules have slightly higher mean 2C19 pIC_{50}s (4.8), followed by bases (4.7), zwitterions (4.5), and acids (4.4) (Figure 10b). Furthermore, while 2C9 and 2C19 are structurally similar, a key observation from this analysis and others is that 2C19 does not have the same affinity for acids as that found for 2C9.[98]

From Figure 10c, one can see that as the clogP increases 2C19 inhibition on average increases.[94] The dependency between 2C19 inhibition and clogP is essentially the same broken down by ionization state, as would be expected given the minor effect of the latter.

2.5.2.4. P450 2D6 Inhibition. Cytochrome P450 2D6 accounts for just 2% of hepatic P450s,[86] however, it is a very important isoform that is implicated in the metabolism of drug molecules such as codeine, ondansetron, and quinidine.[89,94]

From an analysis of the 50886 molecules with measured 2D6 pIC_{50}s one finds that inhibition has a weak but clearly evident parabolic relationship with MWT, similar to that observed for 2C9 (Figure 10d). Similarly, this also suggests steric factors associated with the 2D6 active generally prevent very large or very small molecules from binding optimally.

Basic molecules[99,100] have the highest mean 2D6 pIC_{50}s, followed by zwitterions, neutral molecules, and finally acidic molecules (Figure 10e). This result is in line with X-ray crystallography data, which indicates that an acidic residue (Asp301) in the active site is capable of hydrogen bonding to bound inhibitors or substrates.[101]

Analysis of Figure 10f shows that as lipophilicity increases 2D6 inhibition on average increases, and this difference is not simply due to increases in MWT. The dependency between 2D6 inhibition and clogP is somewhat different when broken down by ionization state, with an increase in clogP having a larger effect on neutral, basic, and zwitterionic molecules compared to that of acids. This would be expected due to the acidic nature of the active site.

2.5.2.5. P450 3A4 Inhibition. The cytochrome P450 3A family accounts for approximately 30% of hepatic P450s, with the 3A4 isoform being the most abundant.[86] Cytochrome P450 3A4 is implicated in the metabolism of drug molecules such as ketoconazole, lidocaine, and erythromycin.[89,94]

From an analysis of 42987 molecules with measured 3A4 pIC_{50}s, one finds that as MWT increases inhibition on average increases (Figure 11a). The MWT dependence of 3A4 is markedly different to the other four CYP isoforms discussed thus far, with no mean drop in potency observed. Surprisingly, the MWT > 700 bin has the same mean pIC_{50} value as the 500–700 bin, indicating the 3A4 active site is considerably larger than those of the other isoforms considered here.[102] Based on an X-ray crystallography study by Williams et al.,[103] this does not appear to be the case. However, a more recent paper by Yano et al.[104] highlighted that while the active site may be the same, the cavity toward the heme is considerably larger than that of other isoforms. Furthermore, studies by Ekroos et al.[105] show that the cavity has the potential to expand considerably on substrate/inhibitor binding.

Ionization state plays a role in determining the extent of inhibition, with neutral molecules having the higher 3A4 pIC_{50}s on average (5.2) than bases or zwitterions (5.1 and 5.0, respectively), which are in turn larger than acids (4.7; Figure 11b). The mean potency of bases is larger than acids and almost equivalent to neutral molecules for a given clogP, suggesting that a specific interaction with an acidic residue in the active site may be possible.

3A4 shows the characteristic dependency on lipophilicity, with the mean 3A4 pIC_{50} increasing as the clogP bin increases, and it is not simply due to increases in MWT (Figure 11c). This result might be expected given the phenylalanine cluster that plays an important role in defining the 3A4 active site and also based on the results from pharmacophoric modeling where multiple hydrophobic features are usual.[14,106]

2.6. Rules of Thumb for a Given Set Molecular Properties. From the results reported herein, one can qualitatively predict the ADMET issues most likely to be experienced for a molecule based on its clogP, MWT, and ionization state, without the need for complex computer simulations. The likelihood of

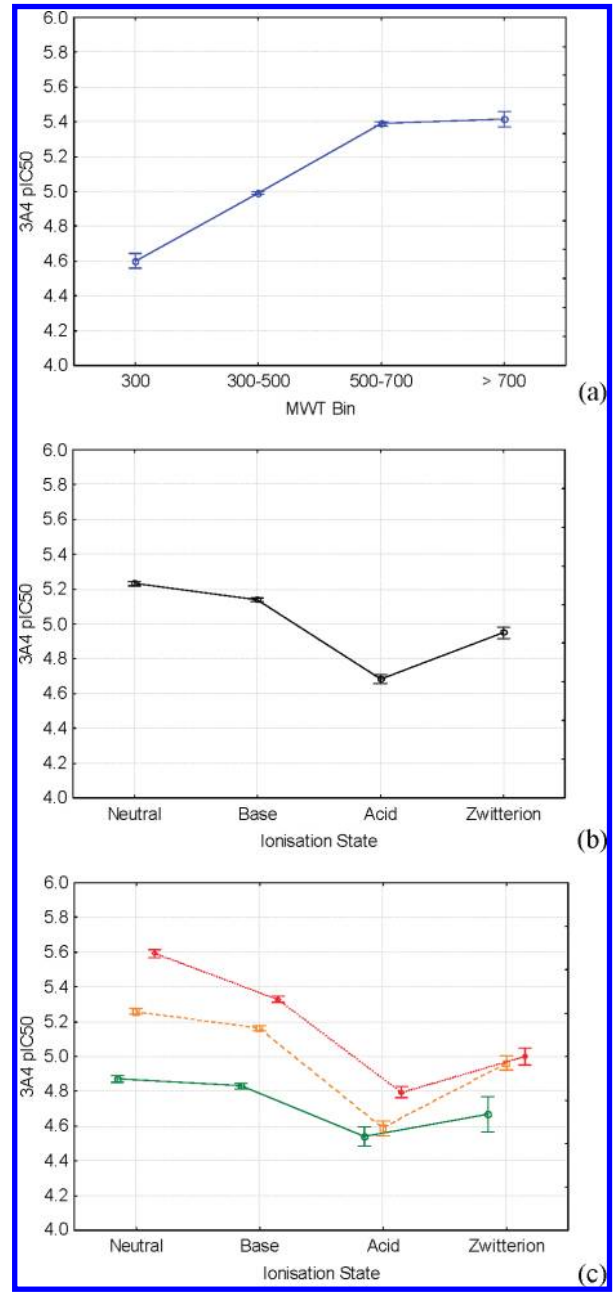

Figure 11. (a−c) Relationship between CYP-3A4 inhibition and MWT (top), ionization state (middle), and a combination of clogP and ionization state (bottom) for 42987 molecules with measured values. The coloring system is the same as that in Figure 2.

a molecule having a particular ADMET parameter above average, average, or less than average is reported in Table 3. Because it is clear that almost all ADMET parameters increase with either increasing MWT and/or clogP, a single combined clogP/MWT category has been used for simplicity. Molecules lie in the more desirable category if both MWT < 400 and clogP < 4, while they are classified as less-desirable should one or more of the parameter lie above the cut-offs.

If we take a hypothetical program series consisting of a neutral chemotype with a low clogP and MWT as an example, one would expect it to have average solubility, higher than average permeability (unless clogP < 3), average protein binding, lower than average hERG inhibition, and higher than average 1A2 inhibition (although the average 1A2 pIC_{50} is just 4.5). If, for example, an increase in solubility and a decrease in CNS penetration were required, and assuming the particular

Table 3. Indication of How Changes in Key Molecular Properties will Affect a Range of ADMET Parameters[a]

neutral molecules	MWT < 400 and clogP < 4	MWT > 400 and/or clogP > 4	
solubility	average	lower	
permeability*	higher	average/higher	
bioavailability	average	lower	
volume of Dist.**	average	average	
plasma protein binding	average	higher	
CNS penetration***	higher/average	average/lower	
brain tissue binding	lower	higher	
P-gp efflux	average	higher/average	
in-vivo clearance	average	average	
hERG Inhibition	lower	lower	
P450 inhibition****	lower 2C9, 2C19, 2D6 & 3A4 inhibition	higher 2C9, 2C19 & 3A4 inhibition	
P450 inhibition****	higher 1A2 inhibition	lower 1A2 inhibition	
P450 inhibition****		average 2D6 inhibition	(a)

basic molecules	MWT < 400 and clogP < 4	MWT > 400 and/or clogP > 4	
solubility	higher/average	lower/average	
permeability*	higher/average	average	
bioavailability	average	lower	
volume of Dist.**	higher/average	higher	
plasma protein binding	lower	average	
CNS penetration***	higher/average	average/lower	
brain tissue binding	lower	higher	
P-gp efflux	average	higher/average	
in-vivo clearance	average	higher/average	
hERG Inhibition	average/higher	higher	
P450 inhibition****	lower 1A2, 2C9, & 2C19 inhibition	lower 1A2 inhibition	
P450 inhibition****	average 2D6 & 3A4 inhibition	average 2C9, 2C19 inhibition	
P450 inhibition****		higher 2D6 & 3A4 inhibition	(b)

acidic molecules	MWT < 400 and clogP < 4	MWT > 400 and/or clogP > 4	
solubility	higher	average/higher	
permeability*	lower	average/lower	
bioavailability	average	average	
volume of Dist.**	lower	lower	
plasma protein binding	average/higher	higher	
CNS penetration***	lower	lower	
brain tissue binding	lower	higher	
P-gp efflux	lower	lower	
in-vivo clearance	lower/average	average	
hERG Inhibition	lower	lower	
P450 inhibition****	lower 1A2, 2C9, 2C19, 2D6 & 3A4 inhibition	lower 1A2, 2C19, 2D6 & 3A4 inhibition	
P450 inhibition****		higher 2C9 inhibition	(c)

(d) zwitterionic molecules	MWT < 400 and clogP < 4	MWT > 400 and/or clogP > 4	
solubility	higher	average/higher	
permeability*	lower	lower/average	
bioavailability	lower	lower	
volume of Dist.**	lower	average/lower	
plasma protein binding	average/lower	higher	
CNS penetration***	average/lower	lower	
brain tissue binding	lower	higher	
P-gp efflux	average	average	
in-vivo clearance	average	average	
hERG Inhibition	lower	average/lower	
P450 inhibition****	lower 1A2, 2C9, 2C19, 2D6 & 3A4 inhibition	lower 1A2, 2C19 & 3A4 inhibition	
P450 inhibition****		average 2C9, 2D6 inhibition	(d)

[a] Expressed relative to the mean value of the data sets. MWT and clogP cut-offs of 400 and 4, respectively, are used. * Optimum clogP bin is 3–5 with respect to permeability. ** Average to high volumes rather than high, low, or average generally considered optimum. *** Low CNS considered optimum, although for targets in the brain, this will be reversed. **** Some isoforms show a nonlinear relationship with clogP and/or MWT. These are guides only. For greater detail, look at the individual ADMET ANOVA graphs found in the text or the tables reported in the Supporting Information.

receptor pharmacophore would allow such a change, one might choose to introduce an acidic group. From Table 3c one would expect that the CNS penetration would decrease on average and solubility would increase on average, as desired. However, undesired effects of this structural change would be that the AM permeability would be expected to decrease on average, as would the volume of distribution, while the plasma protein binding would increase on average.

If we take a second hypothetical program, consisting of a neutral chemotype with a clogP of 4 and a MWT of 400 (Table 3a), where an increase in the potency is required, the MWT would generally be expected to increase during LO.[107] Thus, one would therefore have an expectation that the following ADMET parameters in particular would change; the solubility would decrease, the permeability would decrease, especially if the clogP strayed above 5, and an increase in protein binding and increased CYP inhibition on average would be observed.

It can be seen from the hypothetical examples above that the rather simplistic modeling used here has the advantage of allowing us to make cross comparisons between a large number of ADMET assays. It then becomes easy to assess in a qualitative fashion how changes in the key physicochemical parameters will impact each of the different ADMET parameters in a particular program series. This is a key limitation of multivariate models derived with different 1D, 2D, and 3D molecular descriptors.

This simplicity can be useful in a lead optimization environment where one does not optimize ADMET parameters in isolation. Such simple rules could also be used in the Hit-to-Lead stage to identify the likely ADMET issues of a given lead, allowing resources to be more effectively directed to the areas identified before the molecule enters lead optimization.

3. Conclusions

Described herein are a set of simple, consistent analyses that have been performed on a range of comprehensive ADMET data sets in an effort to improve our general understanding of the main ADMET processes. These rules are consistent with reports in the literature and can be used to supplement the more complex, predictive in-silico models available to us.

The simple rules have been derived using large numbers of measurements from multiple ADMET assays run within GSK. MWT and logP are the two key characteristics that determine ADMET liabilities, with some ADMET parameters depending more on MWT and some on logP. Increasing values in either parameter are generally detrimental to more than one ADMET parameter. One might also expect that the multitude of other pharmacokinetic and toxicity parameters not routinely measured will show a similar dependency. Ionization state can have either a beneficial or a detrimental affect depending on the parameter in question.

4. Experimental Section

4.1. Experimental Data. Experimental data from 15 different ADMET assays run within GSK were extracted from our in-house database. A description of the GSK assays employed here can be found in the following references: Solubility,[111] Permeability,[43] Oral Bioavailablity,[112–114] Volume of Distribution,[112,114–116] Plasma Protein Binding,[117] CNS Penetration,[118] P-gp efflux,[118] Brain Tissue Binding,[118] In-vivo Clearance[112115,116], hERG Inhibition,[119] P450 Inhibition.[120] Approximately 1.0 million measurements in total were available for ~300000 unique molecules. Replicate measurements in a given assay were transformed to a linear scale and averaged (Table 1). The number of molecules measured in each

assay ranges from ~1000 to ~50000, spanning a large range in response values and representing a structurally diverse chemical space.

The data sets were acquired over a period of 1−5 years. Data reported in older variants of a particular assay (e.g., having a different end point detection method or probe) were combined where the interassay correlation was sufficiently strong. Inclusion criteria for P450 data, for example, were given the intra-assay variability of ~2-fold error, older assay variants needed to show interassay variability of ~3-fold to be combined. The final step in the data set definition process was to remove any measurement reported with an adverse comment or modifier (</>/~), giving rise to the number of datapoints reported in the text.

4.2. Molecular Descriptors. Twelve commonly used descriptors were computed using a combination of in-house and commercial software (Table 2),[23] including calculated logP and logD from ACD[122] and additional properties via Daylight.[123] The descriptors computed are quite comprehensive, governing the key bulk characteristics of a molecule. As it is the desire here to formulate a set of simple rules using a consistent set of descriptors across all assays, we therefore require the smallest number that describes the greatest amount of structural variability. This was achieved by building a PCA[18−22] model on ~30000 diverse GSK molecules for which the 12 descriptors in Table 2 had been computed.

As we make changes to the structure of a given chemical in a series through the addition of a variety of substituents at different positions, each of the commonly used interpretable descriptors in Table 2 will change, and these changes are often correlated. Thus, it becomes difficult to fully account for the multidimensional changes in structure. For example, as one increases the number of hydrogen bond acceptors, the clogP will generally decrease and the PSA will increase. This means there is significant redundancy in the descriptors commonly used.

Fortunately, the correlated nature of the structural and electronic changes makes this problem highly amenable to PCA. PCA is a statistical method that can be used to reduce the amount of data to be analyzed by exploiting the correlated nature of the variables within the data set. Linear combinations of the correlated variables are chosen such that the majority of the variance of the original data can be described by a smaller number of orthogonal components.

The PCA model was generated in SIMCA-P10[108] using the default settings.

4.3. Statistical Analyses. ANOVA is the technique used to assess the statistical significance of the relationship between the experimental ADMET parameters and the properties identified from the PCA model. In one-way and factorial ANOVA used here a test is performed to assess the statistical significance of the differences in the means of particular ADMET parameter when broken down into different groups. Results are considered statistical significance here if found above the 99.9% confidence level. This means there is a 1 in 1000 chance that the observed effect is not real.

The three descriptors identified from the PCA model were binned into between three and four separate categories for the analysis, the cut-offs chosen to ensure that a sufficiently large number of observations were present in each bin to make the statistical tests reliable: (a) MWT bin (<300, 300–500, 500–700, and >700), (b) ionization state (neutral, base, acid, and zwitterions), and (c) ACD clogP[109] bin (<3, 3–5, >5). The MWT and clogP bins were chosen to ensure that the observations were relatively evenly distributed.

The ANOVA results were assessed in terms of (a) MWT, (b) ionization state, and (c) clogP and ionization state combined and are presented graphically. The latter combination was investigated as ionization state often affects a particular parameter above the effect on logD. In this way, both affects can be clearly seen. All effects are significant, except where discussed.

To assess whether there is an independent effect from clogP and MWT on a given ADMET parameter, the impact of MWT above any change in clogP has been assessed in all cases using the property binning scheme described and ANOVA.

ANOVA calculations were performed in Statistica 7[110] using the default settings.

Acknowledgment. The author would like to thank Drs. Anne Hersey, Mike Hann, Andrew Leach, and Paul Faulder for carefully reading the manuscript and for making useful suggestions to improve it. The author would also like to thank Drs. Sandeep Modi, Colin Edge, Iain McLay, Andrew Brewster, Paul Smith, Rod Porter, Andy Gribble, Philip Jeffrey, and Peter Lovell for helpful discussions on the subject matter.

Note Added after ASAP Publication. This manuscript was released ASAP on January 31, 2008 with the wrong image for Figure 8. The corrected version was posted on February 7, 2008.

Supporting Information Available: Tables with the ANOVA model results are provided (means, confidence intervals, etc.), a plot illustrating the relationship between pK_a and ionization state, and the scores plot associated with the PCA plot illustrated in Figure 1 are provided. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Grandison, M. K.; Boudinot, F. D. Age-related changes in protein binding of drugs. Implications for therapy. *Clin. Pharmacokinet.* **2000**, *38*, 271–290.
(2) Benet, L. Z.; Hoener, B. A. Changes in plasma protein binding have little clinical relevance. *Clin. Pharmacol. Ther.* **2002**, *71*, 115–121.
(3) Viskin, S. Long QT syndromes and torsade de pointes. *Lancet* **1999**, *354*, 1625–1633.
(4) Rettie, A. E.; Korzekwa, K. R.; Kunze, K.; Lawrence, R. F.; Eddy, A. C.; Aoyama, T.; Gelboin, H. V.; Gonzalez, F. J.; Trager, W. F. Hydroxylation of warfarin by human cDNA-expressed cytochrome P450: a role for P-4502C9 in the etiology of (*S*)-warfarin−drug interactions. *Chem. Res. Toxicol.* **1992**, *5*, 54–59.
(5) Norinder, U.; Bergstrom, C. A. S. Prediction of ADMET properties. *ChemMedChem* **2006**, *1*, 920–937.
(6) Hou, T.; Wang, J.; Zhang, W.; Wang, W.; Xu, X. Recent advances in computational prediction of drug absorption and permeability in drug discovery. *Curr. Med. Chem.* **2006**, *13*, 2653–2667.
(7) Chohan, K. K.; Paine, S. W.; Waters, N. J. Quantitative structure−activity relationships in drug metabolism. *Curr. Top. Med. Chem.* **2006**, *6*, 1569–1578.
(8) De Groot, M. Designing better drugs: Predicting cytochrome P450 metabolism. *Drug Discovery Today* **2006**, *11*, 601–606.
(9) Hansch, D.; Leo, A.; Mekapati, S. B.; Kurup, A. QSAR and ADME. *Bioorg. Med. Chem.* **2004**, *12*, 3391–3400.
(10) Lombardo, F.; Gifford, F.; Shalaeva, M. Y. in-silico ADME prediction: Data, models, facts, and myths. *Mini-Rev. Med. Chem.* **2003**, *3*, 861–875.
(11) Hall, L. H.; Kier, L. B. Electrotopological state indices for atoms types: A Novel combination of electronic, topological, and valence state information. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039–1045.
(12) Burden, F. R. Molecular identification number for substructure searches. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 225–227.
(13) Pearlman, R. S.; Smith, K. M. Novel software tools for chemical diversity. *Perspect. Drug Discovery Des.* **1998**, *9*, 339–353.
(14) deGroot, M.; Ekins, S. Pharmacophore modeling of cytochromes P450. *Adv. Drug Delivery Rev.* **2002**, *54*, 367–383.
(15) Cavalli, A.; Poluzzi, E.; De Ponti, F.; Recanatini, M. Toward a pharmacophore for drugs inducing the long QT syndrome: Insights from a CoMFA study of HERG K+ channel blockers. *J. Med. Chem.* **2002**, *45*, 3844–3853.
(16) 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.
(17) Van de Waterbeemd, H.; Camenisch, G.; Folkers, G.; Chretien, J. R.; Raevsky, O. A. Estimation of blood−brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors. *J. Drug Targeting* **1998**, *6*, 151–165.
(18) Höskuldsson, A. *Prediction Methods in Science and Technology*; Thor Publishing: Copenhagen, Denmark, 1996.
(19) Wold, S.; Geladi, P.; Esbensen, K.; Öhman, J. J. Multi-way principal components-and PLS-analysis. *J. Chemom.* **1987**, *1*, 41–46.

(20) Wold, H. Path models with latent variables: The NIPALS approach. In *Quantitative Sociology: International perspectives on mathematical and statistical model building*; Academic Press: New York, 1975; pp 307–357.
(21) Wold, S.; Albano, C.; Dunn, W. J.; Edlund, U., Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjöström, M. Multivariate data analysis in chemistry. In *Chemometrics: Mathematics and statistics in chemistry*; Kowalski, B. R. Ed.; D. Reidel Publishing Company: Dordrecht, Holland, 1984.
(22) Wold, S.; Eriksson, L.; Sjöström, M. *PLS in Chemistry, Encyclopedia of Computational Chemistry*, Wiley: New York, 2000.
(23) Gleeson, M. P. Plasma protein binding affinity and its relationship to molecular structure: An in-silico analysis. *J. Med. Chem.* **2007**, *50*, 101–112.
(24) An advantage of the ANOVA method used here is that one can directly compare results across all the assays studied in a simple manner, which would be difficult if different descriptors, cut-offs, or modelling methods were employed in each case. However, it should be noted that the independent effect of descriptors, such as clogP and MWT, may not be picked up due to the crude way they are used in this study (small numbers of binned values rather than continuous descriptors). In cases where the effect of each variable is different, this might not be seen at the 99.9% confidence level due to the limitation of the descriptors but also due to low numbers of observation.
(25) Kerns, E. H.; Di, L. Physicochemical profiling: overview of the screens. *Drug Discovery Today* **2004**, *1*, 343–348.
(26) Abraham, M. H.; Le, J. The correlation and prediction of the solubility of molecules in water using an amended solvation energy relationship. *J. Pharm. Sci.* **1999**, *88*, 868–880.
(27) Votano, J. R.; Parham, M.; Hall, L. H.; Kier, L. B. New predictors for several ADME/Tox properties: Aqueous solubility, human oral absorption, and Ames genotoxicity using topological descriptors. *Mol. Diversity* **2004**, *8*, 379–391.
(28) Hansen, N. T.; Kouskoumvekaki, I.; Jørgensen, F. S.; Brunak, S.; Jonsdottir, S. O. Prediction of pH-dependent aqueous solubility of druglike molecules. *J. Chem. Inf. Model.* **2006**, *46*, 2601–2609.
(29) Delaney, J. S. ESOL: Estimating aqueous solubility directly from molecular structure. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1000–1005.
(30) Ran, Y.; Yalkowsky, S. H. Prediction of drug solubility by the general solubility equation (GSE). *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 354–357.
(31) Kansy, M.; Senner, F.; Gubernator, K. Physicochemical high throughput screening: Parallel artificial membrane permeation assay in the description of passive absorption processes. *J. Med. Chem.* **1998**, *41*, 1007–1010.
(32) Cho, M. J.; Thompson, D. P.; Cramer, C. T.; Vidmar, T. J.; Scieszka, J. F. The Madin−Darby canine kidney (MDCK) epithelial cell monolayer as a model cellular transport barrier. *Pharm. Res.* **1989**, *6*, 71–77.
(33) 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.
(34) Camenisch, G.; Alsenz, J.; van de Waterbeemd, H.; Folkers, J. 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.
(35) van de Waterbeemd, H. G.; Camenisch, G.; Folkers, O. A.; Raevsky. Estimation of Caco-2 cell permeability using calculated molecular descriptors. *Quant. Struct.-Act. Relat.* **1996**, *15*, 480–490.
(36) 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.
(37) Tantishaiyakul, V. Prediction of Caco-2 cell permeability using partial least square multivariate analysis. *Pharmazie* **2001**, *56*, 407–411.
(38) Riley, R. J.; Martin, I. J.; Cooper, A. E. The influence of DMPK as an integrated partner in modern drug discovery. *Curr. Drug Metab.* **2002**, *3*, 527–550.
(39) Martinez, M. N.; Amidon, G. L. A mechanistic approach to understanding the factors affecting drug absorption: A review of fundamentals. *J. Clin. Pharmacol.* **2002**, *42*, 620–643.
(40) Turner, J. V.; Glass, B. D.; Agatonovic-Kustrin, S. Prediction of drug bioavailability based on molecular structure. *Anal. Chim. Acta* **2003**, *485*, 89–102.
(41) Yoshida, F.; Topliss, J. G. QSAR Model for Drug Human Oral Bioavailability. *J. Med. Chem.* **2000**, *43*, 2575–2585.
(42) Martin, Y. C. A Bioavailability Score. *J. Med. Chem.* **2005**, *48*, 3164–3170.
(43) Veber, D. F.; Johnson, S. R.; Cheng, H.; 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.
(44) Gleeson, M. P. In silico human and rat V_{ss} quantitative structure−activity relationship models. *J. Med. Chem.* **2006**, *49*, 1953–1963.
(45) Lombardo, F.; Obach, R. S.; Shalaeva, M. Y.†; Gao, F. Prediction of volume of distribution values in humans for neutral and basic drugs using physicochemical measurements and plasma protein binding data. *J. Med. Chem.* **2002**, *45*, 2867–2876.
(46) Ghafourian, T.; Barzegar-Jalali, M.; Hakimiha, N.; Cronin, M. T. D. Quantitative structure−pharmacokinetic relationship modelling: Apparent volume of distribution. *J. Pharm. Pharmacol.* **2004**, *56*, 339–350.
(47) Hitchcock, S. A.; Pennington, L. D. Structure−brain exposure relationships. *J. Med. Chem.* **2006**, *49*, 7559–7583.
(48) Gerebtzoff, G.; Seelig, A. In silico prediction of blood−brain barrier permeation using the calculated molecular cross-sectional area as main parameter. *J. Chem. Inf. Model.* **2006**, *46*, 2638–2650.
(49) Platts, J. A.; Abraham, M. H.; Zhao, Y. H.; Hersey, A.; Ijaz, L.; Butina, D. Correlation and prediction of a large blood-brain distribution data set—an LFER study. *Eur. J. Med. Chem.* **2001**, *36*, 719–730.
(50) Clark, D. E. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 2. Prediction of blood−brain barrier penetration. *J. Pharm. Sci.* **1999**, *88*, 815–821.
(51) Seelig, A. A general pattern for substrate recognition by P-glycoprotein. *Eur. J. Biochem.* **1998**, *251*, 252–261.
(52) Zhang, E. Y.; Phelps, M. A.; Cheng, C.; Ekins, S.; Swaan, P. W. Modeling of active transport systems. *Adv. Drug Delivery Rev.* **2002**, *54*, 329–354.
(53) Gombar, V. K.; Polli, J. W.; Humphreys, J. E.; Wring, S. A.; Serabjit-Singh, C. S. Predicting P-glycoprotein substrates by a quantitative structure−activity relationship model. *J. Pharm. Sci.* **2004**, *93*, 957–968.
(54) Bain, L. J.; LeBlanc, G. A. Interaction of structurally diverse pesticides with the human MDR1 gene product P-glycoprotein. *Toxicol. Appl. Pharmacol.* **1996**, *141*, 288–298.
(55) Litman, T.; Zeuthen, T.; Skovsgaard; T.; Stein, W. D. Structure−activity relationships of P-glycoprotein interacting drugs: Kinetic characterization of their effects on ATPase activity. *Biochim. Biophys. Acta* **1997**, *1361*, 159–168.
(56) Klippert, P.; Borm, P.; Noordhoek, J. Prediction of intestinal firstpass effect of phenacetin in the rat from enzyme kinetic data-correlation with in vivo data using mucosal blood flow. *Biochem. Pharmacol.* **1982**, *31*, 2545–2548.
(57) Oie, S.; Tozer, T. N. Effect of altered plasma protein binding on apparent volume of distribution. *J. Pharm. Sci.* **1979**, *68*, 1203–1205.
(58) Abraham, M. H. Scales of solute hydrogen bonding: Their construction and application to physicochemical and biological processes. *Chem. Soc. Rev.* **1993**, *22*, 73–83.
(59) Davis, A. M.; Riley, R. Impact of physical organic chemistry on the control of drug-like properties. In *Drug design cutting edge*; Flower D. R., Ed.; Royal Society of Chemistry: Cambridge, U.K., 2002; pp 106–123.
(60) Valko, K.; Nunhuck, S.; Bevan, C.; Abraham, M. H.; Reynolds, D. P. Fast gradient HPLC method to determine molecules binding to human serum albumin. Relationships with octanol/water and immobilized artificial membrane lipophilicity. *J. Pharm. Sci.* **2003**, *92* (11), 2236–2248.
(61) Kratochwil, N. A.; Huber, W.; Muller, F.; Kansy, M.; Gerber, P. R. Plasma protein binding of drugs. A new approach. *Biochem. Pharmacol.* **2002**, *64*, 1355–1374.
(62) 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.
(63) Lobell, M.; Sivarajah, V. In silico prediction of aqueous solubility, human plasma protein binding and volume of distribution of molecules from calculated pKa and AlogP98 values. *Mol. Diversity* **2003**, *7*, 69–87.
(64) Yamakazi, K.; Kanaoka, M. Computational prediction of plasma protein binding percent of diverse pharmaceutical molecules. *J. Pharm. Sci.* **2004**, *93*, 1480–1494.
(65) Saiakhov, R. D.; Stefan, L. R.; Klopman, G. Multiple computer-automated structure evaluation model of plasma protein binding affinity of diverse drugs. *Perspect. Drug Discovery Des.* **2000**, *19*, 133–155.
(66) Kalvass, J. C.; Maurer, T. S. Influence of nonspecific brain and plasma binding on CNS exposure: implications for rational drug discovery. *Biopharm. Drug Dispos.* **2002**, *23*, 327–338.
(67) Austin, R. P.; Barton, P.; Mohmed, S.; Riley, R. J. The binding of drugs to hepatocytes and its relationship to physicochemical properties. *Drug Metab. Dispos.* **2005**, *333*, 419–425.

(68) Yokogawa, K. K.; Ishizaki, J.; Ohkuma1, S.; Miyamoto, K. Influence of lipophilicity and lysosomal accumulation on tissue distribution kinetics of basic drugs: A physiologically based pharmacokinetic model. *Methods Find. Exp. Clin. Pharmacol.* **2002**, *24*, 81–93.

(69) Madden, J. C.; Cronin, T. D. Structure-based methods for the prediction of drug metabolism. *Expert Opin. Drug Metab. Toxicol.* **2006**, *2*, 545–557.

(70) Kalgutkar, A. S.; Gardner, I.; Obach, R. S.; Shaffer, C. L.; Callegari, E.; Henne, K. R.; Mutlib, A. E.; Dalvie, D. K.; Lee, J. S.; Nakai, Y.; O'Donnell, J. P.; Boer, J.; Harriman, S. P. A comprehensive listing of bioactivation pathways of organic functional groups. *Curr. Drug Metabol.* **2005**, *6*, 161–225.

(71) Klippert, P.; Borm, P.; Noordhoek, J. Prediction of intestinal first-pass effect of phenacetin in the rat from enzyme kinetic data—Correlation with in vivo data using mucosal blood flow. *Biochem. Pharmacol.* **1982**, *31*, 2545–2548.

(72) Grime, K.; Riley, R. J. The impact of in vitro binding on in vitro−in vivo extrapolations, projections of metabolic clearance, and clinical drug−drug interactions. *Curr. Drug Metab.* **2006**, *7*, 251–264.

(73) Viskin, S.; Long, Q. T. Syndromes and torsade de pointes. *Lancet* **1999**, *354*, 1625–1633.

(74) Witchel, H. J.; Milnes, J. T.; Mitcheson, J. S.; Hancox, J. C. Troubleshooting problems with in vitro screening of drugs for QT interval prolongation using HERG K+ channels expressed in mammalian cell lines and Xenopus oocytes. *J. Pharmacol. Toxicol. Methods* **2002**, *48*, 65–80.

(75) 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.

(76) 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.

(77) Keseru, G. M. Prediction of hERG potassium channel affinity by traditional and hologram QSAR methods. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2773–2775.

(78) Vaz, R. J.; Li, Y.; Rampe, D. Human ether-a-go-go related gene (HERG): A chemist's perspective. *Prog. Med. Chem.* **2005**, *43*, 1–18.

(79) Zhu, B.; Jia, Z. J.; Zhang, P.; Su, T.; Huang, W.; Goldman, E.; Tumas, D.; Kadambi, V.; Eddy, P.; Sinha, U.; Scarborough, R. M.; Song, Y. Inhibitory effect of carboxylic acid group on hERG binding. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5507–5512.

(80) Price, D. A.; Armour, D.; de Groot, M.; Leishman, D.; Napier, C.; Perros, M.; Stammen, B. L.; Anthony Wood, A. Overcoming HERG affinity in the discovery of the CCR5 antagonist maraviroc. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4633–4637.

(81) Rendic, S.; Di Carlo, F. J. D. Human cytochrome P450 enzymes: A status report summarizing their reactions, substrates, inducers, and inhibitors. *Drug Metab. Rev.* **1997**, *29*, 413–580.

(82) Guengerich, F. P. Update information of human P450s. *Drug Metab. Rev.* **2002**, *4*, 7–15.

(83) Lin, J. H.; Lu, A. Y. H. Inhibition and induction of cytochrome P450 and the clinical implications. *Clin. Pharmacokinet.* **1998**, *35*, 361–390.

(84) Bertz, R. J.; Granneman, G. R. Use of in vitro and in vivo data to estimate the likelihood of metabolic pharmacokinetic interactions. *Clin. Pharmacokinet.* **1997**, *32*, 210–258.

(85) Rendic, S.; Di Carlo, F. J. Human cytochrome P450 enzymes: A status report summarizing their reactions, substrates, inducers, and Inhibitors. *Drug Metab. Rev.* **1997**, *29*, 413–580.

(86) 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.

(87) Masimirembwa, C. M.; Thompson, R.; Andersson, T. B. In vitro high throughput screening of compounds for favorable metabolic properties in drug discovery. *Comb. Chem. High Throughput Screening* **2001**, *4*, 245–263.

(88) Lewis, D. F. V. Human cytochromes P450 associated with the phase 1 metabolism of drugs and other xenobiotics: A compilation of substrates and inhibitors of the CYP1, CYP2, and CYP3 families. *Curr. Med. Chem.* **2003**, *10*, 1955–1972.

(89) Lewis, D. F. V. Structural characteristics of human P450s involved in drug metabolism: QSARs and lipophilicity profiles. *Toxicology* **2000**, *144*, 197–203.

(90) Sansen, S.; Yano, J. K.; Reynald, R. L.; Schoch, G. A.; Griffin, K. J.; Stout, D. D.; Johnson, E. F. Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2. *J. Biol. Chem.* **2007**, *282* (19), 14348–14355.

(91) Burton, J.; Ijjaali, I.; Barberan, O.; Petitet, F.; Vercauteren, D. P.; Michel, A. Recursive partitioning for the prediction of cytochromes P450 2D6 and 1A2 inhibition: Importance of the quality of the dataset. *J. Med. Chem.* **2006**, *49*, 6231–6240.

(92) Chohan, K. K.; Paine, S. W.; Mistry, J.; Barton, P.; Davis, A. M. A rapid computational filter for cytochrome P450 1A2 inhibition potential of molecule libraries. *J. Med. Chem.* **2005**, *48*, 5154–5161.

(93) Korhonen, L. E.; Rahnasto, M.; Maehoenen, N. J.; Wittekindt, C.; Poso, A.; Juvonen, R. O.; Raunio, H. Predictive three-dimensional quantitative structure−activity relationship of cytochrome P450 1A2 inhibitors. *J. Med. Chem.* **2005**, *48*, 3808–3815.

(94) Lewis, F. V.; Modi, S.; Dickins, M. Quantitative structure−activity relationships (QSARs) within substrates of human cytochrome P450 involved in drug metabolism. *Drug Metab. Drug Rev.* **2001**, *18*, 221–242.

(95) Williams, P. A.; Cosme, J.; Ward, A.; Angove, H. C.; Vinkovic, D. M.; Jhoti, H. Crystal structure of humancytochrome P450 2C9 with bound warfarin. *Nature* **2003**, *424*, 464–468.

(96) Wester, M. R.; Yano, J. K.; Schoch, G. A.; Yang, C.; Griffin, K. J.; Stout, D. D.; Johnson, E. F. The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-Å resolution. *J. Biol. Chem.* **2004**, *279*, 563035637.

(97) Niwa, N.; Kageyama, A.; Kishimoto, K.; Yabusaki, Y.; Ishibashi, F.; Katagiri, M. Amino acid residues affecting the activities of human cytochrome P450 2C9 and 2C19. *Drug Metab. Dispos.* **2002**, *30*, 931–936.

(98) Lewis, D. F. On the recognition of mammalian microsomal cytochrome P450 substrates and their characteristics towards the prediction of human P450 substrate specificity and metabolism. *Biochem. Pharmacol.* **2000**, *60*, 293–306.

(99) Snyder, R.; Sangar, R.; Wang, J.; Ekins, S. Three-dimensional quantitative structure−activity relationship for Cyp2D6 substrates. *Quant. Struct.-Act. Relat.* **2002**, *21*, 357–368.

(100) Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J.; Wikel, J. H.; Wrighton, S. A. Three and four dimensional-quantitative structure−activity relationship (3D/4D-QSAR) analyses of CYP2D6 inhibitors. *Pharmacogenetics* **1999**, *9*, 477–489.

(101) Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R.; Oxbrow, A. K.; Lewis, C. J.; Tennant, M. G.; Modi, S.; Eggleston, D. S.; Chenery, R. J.; Bridges, A. M. Crystal structure of human cytochrome P450 2D6. *J. Biol. Chem.* **2006**, *281*, 7614–7622.

(102) Rendic, S.; Di Carlo, F. J. Human cytochrome P450 enzymes: A status report summarizing their reactions, substrates, inducers, and inhibitors. *Drug Metab. Rev.* **1997**, *29*, 413–580.

(103) Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **2004**, *305*, 683–686.

(104) Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05 Å resolution. *J. Biol. Chem.* **2004**, *279*, 38091–38094.

(105) Ekroos, M.; Sjogren, T. Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13682–13687.

(106) Ekins, S.; Stresser, D. M. ; Williams, A. J. In vitro and pharmacophore insights into CYP3A enzymes. *Trends Pharmacol. Sci.* **2003**, *24*, 161–166.

(107) Hadjuk, P. J. Fragment-based drug design: How big is too big. *J. Med. Chem.* **2006**, *49*, 6972–6976.

(108) *SIMCA-P 10*, Umetrics: Tvistevägen 48, Box 7960, SE-907 19 Umeå, Sweden.

(109) Advanced Chemistry Development, Inc., 110 Yonge Street, 14th floor, Toronto, Ontario, M5C 1T4, Canada, www.acdlabs.com.

(110) Statistica System Reference, Statsoft Inc., 2300 East Tulsa, OK 74104, www.statsoft.com.

(111) Pan, L.; Ho, G.; Tsutsui, K.; Takahashi, L. Comparison of chromatographic and spectroscopic methods used to rank compounds for aqueous solubility. *J. Pharm. Sci.* **2001**, *90*, 521–529.

(112) All animal studies were approved by an internal ethical review committee and performed in accordance with the UK Animals (Scientific Procedures) Act 1986 and "Principles of Laboratory Animal Care" (NIH Publication #85-23, rev. 1985). Standard methodologies were used to determine basic pharmacokinetic parameters following oral or intravenous dosing (e.g., clearance (blood or plasma), volume of distribution, terminal half-life, and oral bioavailability). Serial blood (or plasma) samples were obtained and analysed for parent compound concentrations using LC-MS/MS methodologies. Doses and formulations were selected on the basis of compound solubility and pharmacological activity. To minimize the impact of the individual study design, such as dosing vehicle, amount, and so on, only discrete results where the oral dose was <10 mg/kg in a standard formulation of polyethylene glycol (PEG)

were used. For the purpose of illustration, a number of PK studies by GSK scientists are referenced below (refs 113118).

(113) Watson, N. S.; Brown, D.; Campbell, M.; Chan, C.; Chaudry, L.; Convery, M. A.; Fenwick, R.; Hamblin, J. N.; Haslam, C.; Kelly, H. A.; King, N. P.; Kurtis, C. L.; Leach, A. R.; Manchee, G. R.; Mason, A. M.; Mitchell, C.; Patel, C.; Patel, V. K.; Senger, S.; Shah, G. P.; Weston, H. E.; Whitworth, C.; Young, R. J. Design and synthesis of orally active pyrrolidin-2-one-based factor Xa inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3784–3788.

(114) Austin, N. E.; Baldwin, S. J.; Cutler, L.; Deeks, N.; Kelly, P. J.; Nash, M.; Shardlow, C. E.; Stemp, G.; Thewlis, K.; Ayrton, A.; Jeffrey, P. Pharmacokinetics of the novel, high-affinity, and selective dopamine D3 receptor antagonist SB-277011 in rat, dog, and monkey: In vitro/in vivo correlation and the role of aldehyde oxidase. *Xenobiotica* **2001**, *31*, 677–686.

(115) Forbes, I. T.; Douglas, S.; Gribble, A. D.; Ife, R. J.; Lightfoot, A. P.; Garner, A. E.; Riley, G. J.; Jeffrey, P.; Stevens, A. J.; Stean, T. O.; Thomas, D. R. SB-656104-A: A novel 5-HT7 receptor antagonist with improved In vivo properties. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 3341–3344.

(116) Hagan, J. J.; Price, G. W.; Jeffrey, P.; Deeks, N. J.; Stean, T.; Piper, D.; Smith, M. I.; Upton, N.; Medhurst, A. D.; Middlemiss, D. N.; Riley, G. J.; Lovell, P. J.; Bromidge, S. M.; Thomas, D. R. Characterization of SB-269970-A, a selective 5-HT7 receptor antagonist. *Br. J. Pharmacol.* **2000**, *130*, 539–548.

(117) Plasma protein binding data was measured in the same experimental assay as that described in ref 118. The key difference was that the blood sample was centrifuged and the analysis was performed on the plasma sample only.

(118) Summerfield, S. G.; Stevens, A. J.; Cutler, L.; Carmen-Osuna, M. D.; Hammond, B.; Tang, S.; Hersey, A.; Spalding, D. J.; Jeffrey, P. Improving the in vitro prediction of in vivo central nervous system penetration: Integrating permeability, P-glycoprotein efflux, and free fractions in blood and brain. *J. Pharmacol. Exp. Ther.* **2006**, *316*, 1282–1290.

(119) The hERG inhibitory potential was determined using an SPA assay (Scintillation Proximity Assay) based on the CHO K1 hERG cell line using the dofetilide ^{3}H radioligand. To determine the IC_{50} of a compound, 11 different concentrations ranging between 0.1 to 100 μM. The resulting dofetilide ^{3}H response was determined using a Wallac 1430 ViewLux microplate imager. Dose−response curves were fitted using Grafit.

(120) The cytochrome P450 inhibitory potential was determined against CYP1A2, 2C9, 2C19, 2D6, and 3A4 using bactosomes expressing the appropriate cytochrome P450 (Cypex Ltd, Dundee, U.K.). The following probes were used: (3-butyryl-7-methoxycoumarin 7-ethoxyresorufin (1A2), 7-methoxy-4-trifluoromethylcoumarin-3-acetic acid (2C9), 3-butyryl-7-methoxycoumarin (2C19), 4-methylaminomethyl-7-methoxycoumarin (2D6), and diethoxyfluorescein (3A4). The inhibitory activity was determined by a fluorescent plate reader methodology (Tecan Safire II (Tecan, Zurich, Switzerland)), similar to that described by Crespi et al. (ref 121) using a seven-point IC_{50} curve. Dose−response curves were subsequently fitted using Grafit.

(121) Crespi, C. L.; Miller, V. P.; Penman, B. W. Microtiter plate assays for inhibition of human drug-metabolizing cytochromes P450. *Anal. Biochem.* **1997**, *248*, 188–190.

(122) Advanced Chemistry Development, Inc., 110 Yonge Street, 14th Floor, Toronto, Ontario, M5C 1T4, Canada, http://www.acdlabs.com.

(123) Daylight Chemical Information Systems, Inc., 120 Vantis, Suite 550, Aliso Viejo, CA 92656, U.S.A. http://www.daylight.com.

JM701122Q