Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry Letters 17 (2007) 1759–1764
Bioorganic & Medicinal Chemistry Letters

A quantitative assessment of hERG liability as a function of lipophilicity

Michael J. Waring* and Craig Johnstone

AstraZeneca, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK

Received 26 September 2006; revised 15 December 2006; accepted 16 December 2006
Available online 22 December 2006

Abstract—The impact of lipophilicity as a factor contributing to hERG potency is assessed for a large dataset of compounds of differing ionisation type. This dataset is derived from compounds tested in the Ionworks™-based in vitro electrophysiology hERG assay at AstraZeneca. Using logistic regression, a quantification of the risk associated with increasing lipophilicity is presented. The anticipated differences between acidic, basic and neutral compounds are apparent in the data but lipophilicity is shown to be a stronger driver for hERG potency than might have been expected. Simple rules defining target lipophilicity values for minimizing hERG liability are derived.

The interaction of pharmaceutical agents with the human ether-a-go-go related gene (hERG)-encoded cardiac potassium channel represents a significant safety concern.[1] Inhibition of the hERG channel can cause delayed ventricular cell repolarisation (seen on the electrocardiogram as prolongation of the QT interval). In a very small percentage of people, this is associated with a potentially fatal cardiac arrhythmia called Torsades de Pointes. This has resulted in a number of drugs being withdrawn from the market and represents a significant regulatory hurdle for new molecular entities in development.[2]

Due to the promiscuous nature of this interaction, the in vitro screening of many compounds against hERG is now carried out at an early stage in drug discovery projects. In addition, a number of computational methods, both pharmacophore and descriptor based, that would enable a prediction of hERG liability have been reported.[3] Whilst these methods all differ slightly in their source dataset and the level of contribution from various factors, the increased liability associated with basic, lipophilic compounds is suggested strongly in all of these studies. We were interested in the quantification of these factors in order to allow evaluation of the hERG liability of individual compounds or particular series prior to synthesis.

At AstraZeneca, a large dataset has emerged on a diverse range of compounds spanning a number of projects, compound classes and target disease areas. This screening has been based on the Ionworks™ whole-cell electrophysiology-based system.[4] This dataset, coupled with octanol: water distribution coefficients ($\log D$), measured at pH 7.4, and a classification of compounds as acidic, basic, neutral and Zwitterionic, was used to assess the hERG liability associated with these factors.

At the time of analysis, data were available for 7685 compounds. Of the total set, 1211 compounds had $\log D$ values measured at pH 7.4. Calculated $\log D$ values were generated for the total dataset and the calculated data validated against the measured data where available. A number of methods for this purpose are available,[5] but the most accurate method evaluated was the in-house AZlog D algorithm.[6] This is trained directly on in-house, measured $\log D_{7.4}$ data and it predicted the measured data with a root mean square error (RMSE) of 0.57 and the residuals normally distributed about the zero-point for compounds that were not in the original training set (see Supporting information). The compounds were classified according to their anticipated ionisation state (acidic, basic, neutral and Zwitterionic) based on SMILES arbitrary target specification (SMARTS) codes for those groups believed to be ionised at pH 7.4. A set of example substructures illustrating groups considered

Keywords: hERG; Lipophilicity.
* Corresponding author. Tel.: +44 1625 230942; fax: +44 1625 516667; e-mail: mike.j.waring@astrazeneca.com

0960-894X/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2006.12.061

to be acidic and basic is provided in the Supporting information. The relative populations and mean values of pIC_{50}, AZlog D and clog P of these four categories are given in Table 1.

Of course, for ionised compounds, the log D value is a consequence of the partitioning of the neutral form of the compound (log P) and the proportion of the compound which is ionised at pH 7.4. This complicates, to a degree, the use of log D to quantify the intrinsic lipophilicity of these classes. The use of a calculated log P value was also considered. The commercial clog P method[7] was found to handle the compounds in this dataset less well than the AZlog D method; predicting the measured log D values of the neutral compounds with an RMSE of 0.91 and evidence of systematic bias in the prediction (see Supporting information).

In order to maximize the hERG dataset, not to constrain it to those compounds for which log D has been measured and, more importantly, to assess the potential for these analyses to predict future compounds, subsequent analyses have been carried out using the calculated AZlog D and clog P values. Given the differing performance of the two models, it is likely that the more rigorous conclusions should be drawn from the analyses using the AZlog D data. However, since the clog P method is more widely available, there is added value in including complementary analyses using this descriptor.

Table 1. Distributions and mean values of hERG pIC_{50}, AZlog D and clog P of ionisation types within the dataset (standard deviations are shown in parentheses)

	Acid	Base	Neutral	Zwitterion
n	350	4302	2598	435
Mean hERG pIC_{50}	3.7 (0.53)	5.2 (0.89)	4.5 (0.70)	4.4 (0.80)
Mean AZlog D	0.71 (1.4)	2.5 (0.89)	2.9 (0.86)	1.5 (1.2)
Mean clog P	3.1 (2.2)	3.6 (1.4)	3.2 (1.6)	4.4 (0.80)

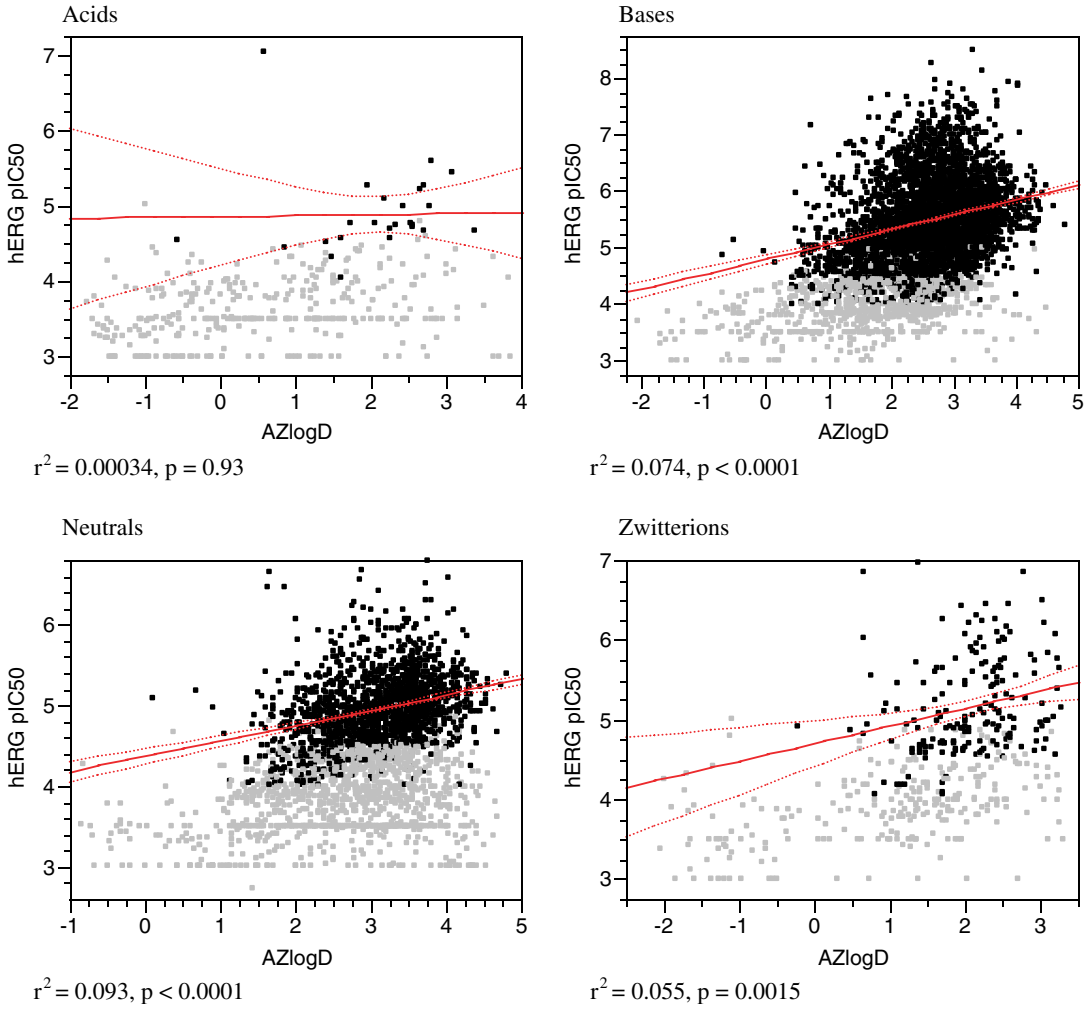

Figure 1. Relationship between AZlog D and hERG pIC_{50} for acids, bases, neutrals and Zwitterions. Compounds with quoted pIC_{50} values above the top concentration in the assay (shown in grey) are excluded from the fit.

The dataset is populated for the most part by compounds from drug discovery projects. Two hundred and fifty-two individual projects are represented within the set. Some projects are represented to a greater extent than others but none represents a significant fraction of the total (the largest contributor represents 3.3% of the data).

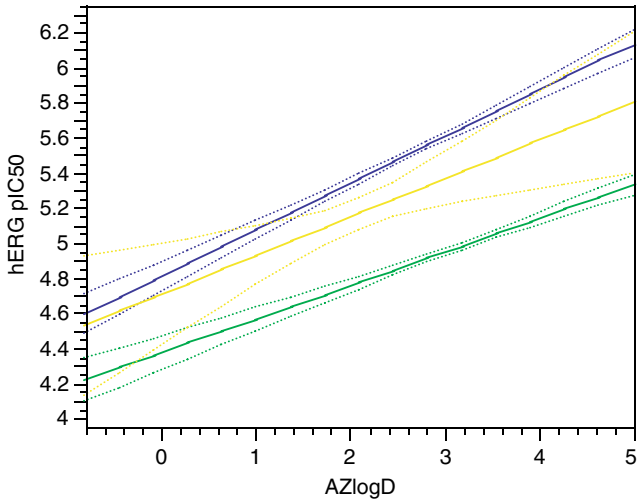

Figure 2. Superimposed lines of best fit and 95% confidence curves for bases (blue), neutrals (green) and Zwitterions (yellow).

The relationship between hERG potency and lipophilicity can be shown for each of the four ionisation categories by plotting AZlog*D* against hERG pIC_{50} (Fig. 1).

As would be anticipated for a diverse set of compounds, there is a large spread in the data and the correlations are weak. However, the overall trend for increasing potential for hERG potency with increasing AZlog*D* is apparent in the lines of best fit for all but the acids, and the *p* values indicate that these relationships are significant. Superimposing the lines of best fit for basic and neutral compounds (Fig. 2) shows that bases have a greater inherent liability as anticipated; the Zwitterions fall somewhere in between. The confidence curves indicate that this difference is significant at the 95% level. It is difficult to compare accurately the relative importance of a continuous variable such as log*D* with a classification such as acidic/basic, but it would appear from these data that lipophilicity is a significant contributor to hERG potency and is of similar importance to ionisation class.

Typically, a drug candidate is required to have a hERG IC_{50} that is above a threshold value. Studies conducting

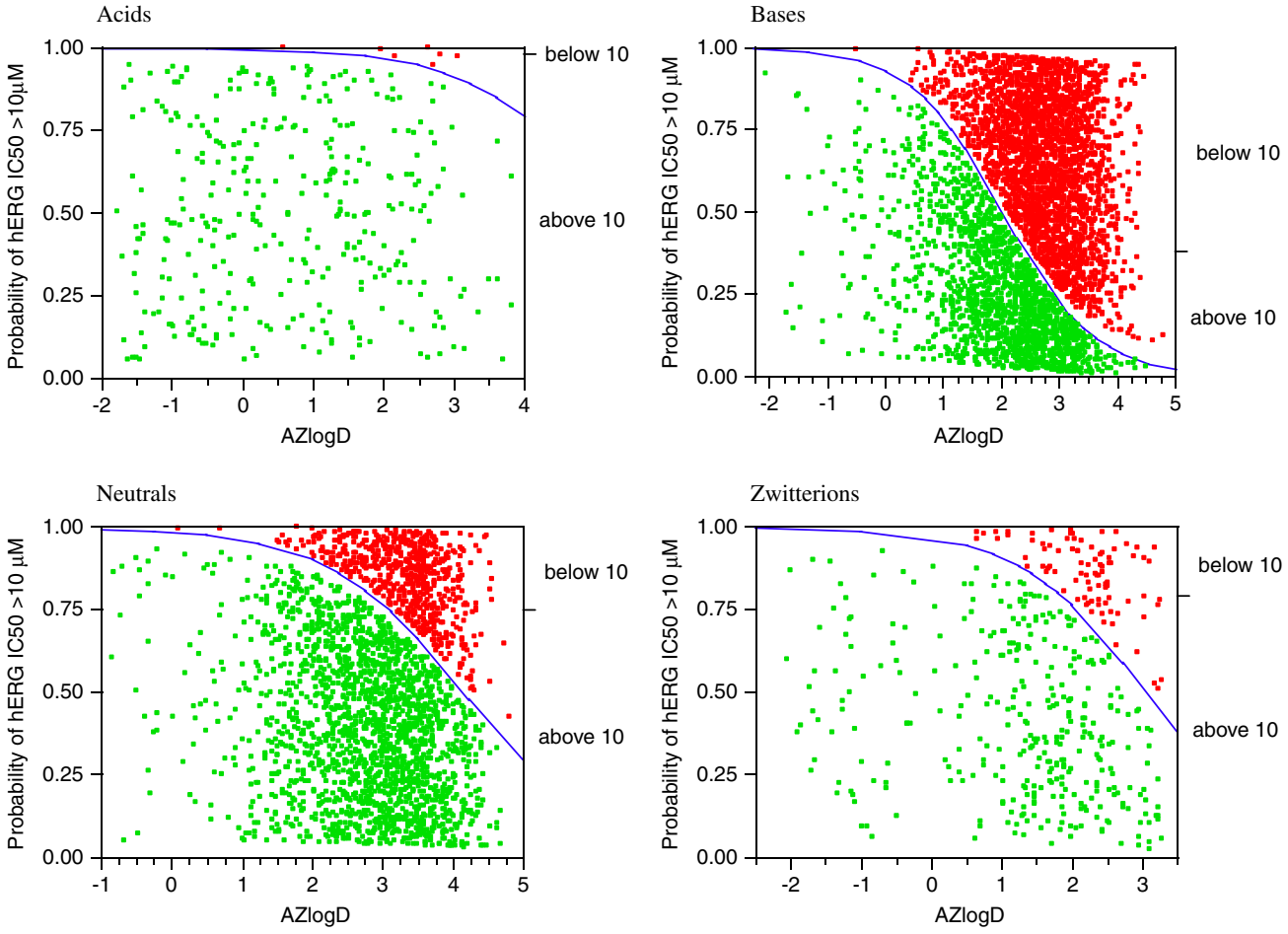

Figure 3. Logistic regressions showing how the probability of a compound achieving a hERG IC_{50} of >10 µM changes with AZlog*D* for each ionisation class. Those compounds with IC_{50} values above 10 µM are shown in green; those below 10 µM are in red.

analyses of the literature suggest that this can be based on the margin between hERG IC_{50} and the maximal free drug concentration required for clinical efficacy[8] or, more crudely, a margin to a primary target potency. In our experience, for typical projects, for a compound to satisfy such criteria, it is usually required to have an absolute hERG IC_{50} of greater than 10 μM. For this reason, we were interested in examining how the probability of meeting such a criterion was influenced by lipophilicity and ionisation.

The compounds were divided into two categories, those with IC_{50} values below 10 μM and those with values above. Logistic regression analyses were carried out to quantify how the likelihood of a compound achieving an IC_{50} of greater than 10 μM diminishes with increasing AZlog *D* (Fig. 3) for each of the four classes.

Table 2. The probability of a compound achieving a hERG IC_{50} of greater than 10 μM for a given log *D*

AZlog *D*	Probability of >10 μM			
	Acids	Bases	Neutrals	Zwitterions
2	0.97	0.50	0.90	0.76
3	0.92	0.24	0.76	0.53
4	—	0.07	0.55	—

The logistic regression technique,[9] although not widely used in medicinal chemistry analysis, is ideally suited to the question at hand. It is a statistical regression model which assesses how relative populations of a series of categories (in this case hERG IC_{50} values above or below 10 μM) change with a continuous variable (AZlog *D* in this case). For such a large, structurally diverse dataset, this can be interpreted as the probability of a future compound of the same value of log *D* having a hERG IC_{50} above the 10 μM threshold.

As expected, the differing liability of the four classes is apparent in these plots, but the increasing liability of all four classes with increasing lipophilicity can be seen clearly. The probabilities of compounds of given AZlog *D* achieving an IC_{50} of greater than 10 μM are summarised (Table 2). Most markedly, the likelihood of achieving this with a basic compound appears very low. Even for non-lipophilic (log *D* = 2) bases, half of these would be anticipated to be unacceptable against this criterion.

As stated previously, complementary analyses were carried out using the commercially available clog *P* method. Whilst this method appears less accurate than the AZlog *D* algorithm, its wider availability makes its con-

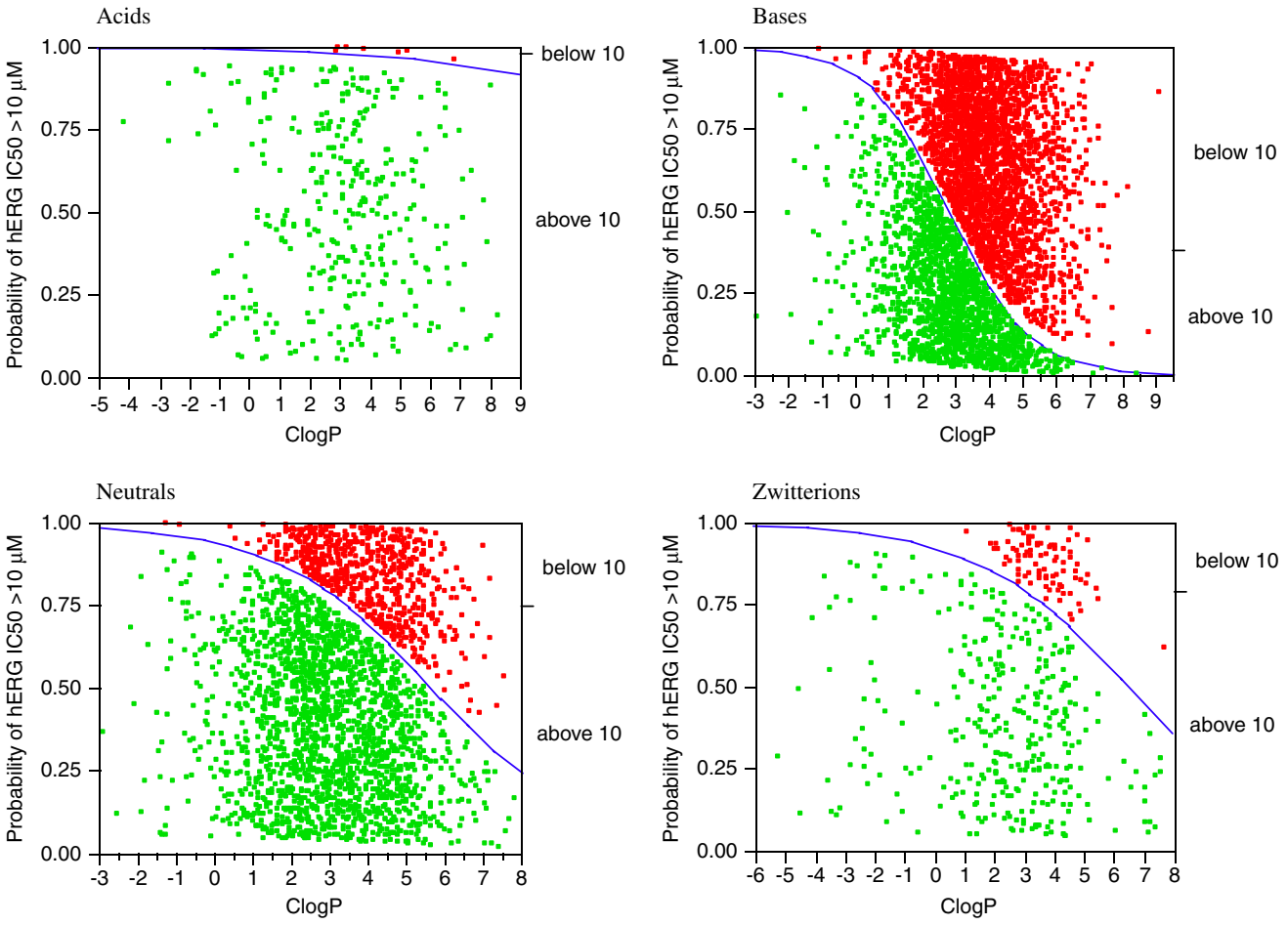

Figure 4. Logistic regressions showing how the probability of a compound achieving a hERG IC_{50} of >10 μM changes with clog *P* for each ionisation class. Those compounds with IC_{50} values above 10 μM are shown in green; those below 10 μM are in red.

sideration valid. In addition, it might be anticipated that the differences between ionisation classes may appear different when considering log *P* as opposed to log *D*.

Logistic regressions against the 10 μM threshold show that a dependence on clog *P* exists within each series (Fig. 4). The same difference between the ionisation types is apparent. There is a difference in the magnitude of the probability when comparing the clog *P* equivalent to the log *D* value reported previously (Table 3). This is likely to be due to the difference between log *D* and log *P* for ionised compounds but could also be attributed, in part, to the systematic error in the clog *P* calculation.

Since the hERG potencies in this analysis are generated in a cellular system, it might be argued that an element of these relationships is due to changes in permeability of the compounds rather than their true activity at the hERG channel; cellular permeability is well established to be related to lipophilicity.[10] This can be addressed by considering the relationship between the Ionworks™ cellular assay and the hERG radioligand binding assay.[11] Whilst discrepancies in the relationship between these two could be ascribed to poor cellular permeability, they could equally be due to other differences between the two assays including differences in the conformation of the expressed hERG channel; the latter is highly likely to be a factor since in the cellular system the hERG channel is membrane bound. However, should there appear to be a bias towards compounds showing increases in potency in the binding assay relative to the cellular system or should the discrepancy between the two appear related to lipophilicity then this might raise concerns that the potency—lipophilicity relationship is artefactual. There are 543 compounds within the dataset that have hERG binding measurements, of these 510 are basic. For these compounds, the pIC_{50} measured in the binding assay was compared to that measured in cells (Fig. 5). Overall, the two IC_{50} values compare reasonably well (RMSE 0.59) and the average difference is not significantly different from zero. Moreover, there is no difference between the RMSE values with changing AZlog *D*.

Carrying out the analogous logistic regressions with the binding assay dataset revealed very similar trends to those observed in the cellular system (Table 4). Given that these are not all the same compounds as measured in the cellular assay, it is likely that the small differences in the results are ascribable to different structures and sample sizes. This suggests that these relationships are not due to differences in permeability of the compounds across the global dataset. However, it may still be true of isolated compounds or sub-series within it.

Table 3. The probability of a compound achieving a hERG IC_{50} of greater than 10 μM for a given clog *P*

clog *P*	Probability of >10 μM			
	Acids	Bases	Neutrals	Zwitterions
2	0.99	0.67	0.85	0.85
3	0.98	0.47	0.79	0.79
4	0.98	0.27	0.67	0.73

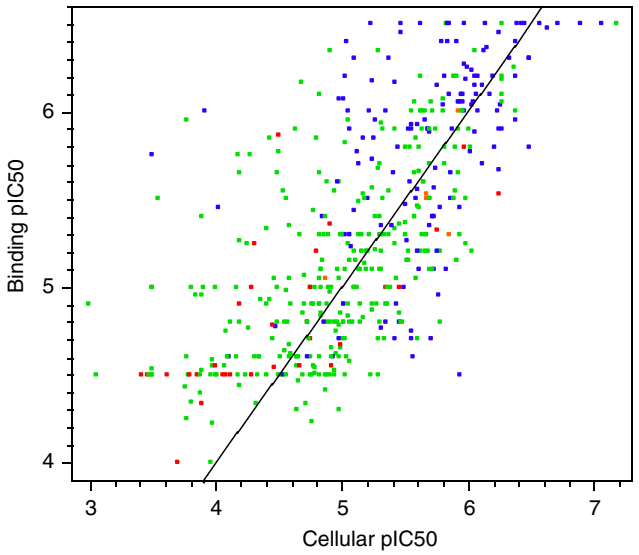

Figure 5. Relationship between hERG binding and cellular pIC_{50} values for a subset of basic compounds. The line shows the 1:1 (y = x) line the overall RMSE around this line is 0.58. The compounds are categorized according to their AZlog *D* values: <1.5 in red, RMSE 0.60; 1.5–3 in green, RMSE 0.55; >3 in blue, RMSE 0.59.

Table 4. The probability of a compound achieving a hERG IC_{50} of greater than 10 μM for a given AZlog *D* in the binding assay

AZlog *D*	Probability of >10 μM			
	Acids (*n* = 126)	Bases (*n* = 2289)	Neutrals (*n* = 292)	Zwitterions (*n* = 65)
2	0.98	0.59	0.98	0.72
3	0.98	0.26	0.93	0.47
4	—	0.08	0.69	—

An alternative way of reading the logistic plots is to use them to define a target log *D* or clog *P* value to use as an upper limit so as to achieve a minimum proportion of compounds being >10 μM. The limits that would be anticipated to lead to >70% of compounds satisfying this criterion are shown for each class (Table 5).

Since the original analyses were carried out, a further 5748 compounds have been tested in the Ionworks™ assay. The data for these compounds could be considered as an independent test of the predictive nature of these rules. The actual percentages of subsequent compounds with hERG IC_{50} > 10 μM at these cut-off values were calculated (Table 6). It is clear that these rules are predicting well the observed attrition for bases and neutrals; this is very close to the predicted 70% figure in all cases. The Zwitterions appear to be handled slightly less well although this may be due, at least in part, to a smaller sample size.

Table 5. Target upper limits of log *D* and clog *P* to ensure >70% of compounds achieve a hERG IC_{50} of greater than 10 μM

	Acids	Bases	Neutrals	Zwitterions
log *D*	>4	1.4	3.3	2.3
clog *P*	>9	1.9	4.0	4.4

Table 6. Test set compounds; actual percentages of compounds with hERG IC_{50} of greater than 10 μM for the values of AZlog *D* and clog *P*, anticipated to give 70% attrition

Bases (n = 2935)		Neutrals (n = 2319)		Zwitterions (n = 243)	
AZlog *D* 1.4	clog *P* 1.9	AZlog *D* 3.3	clog *P* 4	AZlog *D* 2.3	clog *P* 4.4
75%	72%	73%	71%	80%	86%

As anticipated, it is clear that basic compounds are more likely to have hERG activity than neutral compounds which are in turn more liable than acids. However, these analyses suggest that lipophilicity is a contributor comparable in magnitude to ionisation. Targeting lipophilicities below those defined in Table 5 is likely to provide the greatest likelihood of avoiding hERG issues within chemical series. Specific examples of such strategies have been described in the literature.[12] Based on this it has been proposed that lead molecules with clog *P* > 3 can be optimized by reducing lipophilicity should a correlation between clog *P* and hERG be established. Should compounds have clog *P* values <3 and still exhibit hERG potency then structural modification is required. The findings described here are broadly consistent with these observations although it is likely that different limits should be appropriate for different ionisation classes. The low probability of lipophilic bases achieving acceptable hERG potencies might be considered prohibitive in the environment of increasingly stringent regulatory views on hERG and QT. Targets requiring such a motif for primary potency may be required to adopt radical or innovative strategies to avoid the problem. Moreover, the desirability of targeting reduced lipophilicity in searching for drug candidates[13] is reinforced and, perhaps, further rationalised by these findings.

These analyses should be useful in allowing the early assessment of the risk of a new chemical series based on ionisation state and lipophilicity as well as providing a potential strategy for reducing hERG potency within a more established project.

Acknowledgments

The authors thank B. G. Small, M. H. Bridgland-Taylor, A. J. Woods and A. Harmer for generating the IonWorks™ data that form the basis for this work and Drs. K. G. Pike, D. McKerrecher and C. E. Pollard for their helpful discussions and assistance with the compilation of this manuscript.

Supplementary data

Plots validating the AZlog *D* and clog *P* algorithms against measured data, a set of example substructures representing acidic and basic groups and a representative portion of the dataset are available. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2006.12.061.

References and notes

1. (a) Sanguinetti, M. C.; Jiang, C.; Curran, M. E.; Keating, M. T. *Cell* **1995**, *81*, 299; (b) De Ponti, F.; Poluzzi, E.; Cavalli, A.; Recanatini, M.; Montanaro, N. *Drug Saf.* **2002**, *25*, 263; (c) Mitcheson, J. S.; Perry, M. D. *Curr. Opin. Drug Discov. Devel.* **2003**, *6*, 667; (d) Pearlstein, R.; Vaz, R.; Rampe, D. *J. Med. Chem.* **2003**, *46*, 2017.
2. Shah, R. R. *Drug Saf.* **2005**, *28*, 115.
3. (a) Cavalli, A.; Poluzzi, E.; De Ponti, F.; Recanatini, M. *J. Med. Chem.* **2002**, *45*, 3844; (b) Cianchetta, G.; Li, Y.; Kang, J.; Rampe, D.; Fravolini, A.; Cruciani, G.; Vaz, R. J. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3637; (c) Roche, O.; Trube, G.; Zuegge, J.; Pfimlin, P.; Alanine, A.; Schneider, G. *ChemBioChem* **2002**, *3*, 455; (d) Tobita, M.; Nishikawa, T.; Nagashima, R. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2886; (e) In The 14th *European Symposium on Quantitative Structure–Activity Relationships*, Buyck, C., Tollenaere, J., Engels, M., De Clerck, F., Eds. Bournemouth, UK, 2002; (f) Aronov, A. M. *Drug Discov. Today* **2005**, *10*, 149; (g) In *The 2nd Anglo-Swedish Medicinal Chemistry Conference*, Springthorpe, B., Ed. Åre, Sweden, 2005; (h) Farid, R.; Day, T.; Friesner, R. A.; Pearlstein, R. A. *Bioorg. Med. Chem.* **2006**, *14*, 3160.
4. (a) Kiss, L.; Bennett, P. B.; Uebele, V. N.; Koblan, K. S.; Kane, S. A.; Neagle, B.; Schroeder, K. *Assay Drug Dev. Technol.* **2003**, *1*, 127; (b) Bridgland-Taylor, M. H.; Hargreaves, A. C.; Easter, A.; Orme, A.; Harmer, A.; Henthorn, D. C.; Ding, M.; Davis, A.; Small, B. G.; Heapy, C. G.; Abi-Gerges, N.; Paulsson, F.; Jacobson, I.; Schroeder, K.; Neagle, B.; Albertson, N.; Hammond, T. G.; Sullivan, M.; Sullivan, E.; Valentin, J.-P.; Pollard, C. E. *J. Pharmacol. Toxicol. Methods* **2006**, *54*, 189.
5. Livingstone, D. *J. Curr. Top. Med. Chem.* **2003**, *3*, 1171.
6. Bruneau, P.; McElroy, N. R. *J. Chem. Inf. Model.* **2006**, *46*, 1379.
7. Using clogP version 4.71, http://www.daylight.com.
8. (a) Webster, R.; Leishman, D.; Walker, D. *Curr. Opin. Drug Discov. Devel.* **2002**, *5*, 116; (b) Redfern, W. S.; Carlsson, L.; Davis, A. S.; Lynch, W. G.; MacKenzie, I.; Palethorpe, S.; Siegl, P. K. S.; Strang, I.; Sullivan, A. T.; Wallis, R.; Camm, A. J.; Hammond, T. G. *Cardiovasc. Res.* **2003**, *58*, 32; (c) De Bruin, M. L.; Pettersson, L. M.; Meyboom, R. H. B.; Hoes, A. W.; Leufkens, H. G. M. *Eur. Heart J.* **2005**, *26*, 590.
9. Hosmer, D. W.; Lemeshow, S. *Applied logistic regression*, 2nd ed.; Wiley: New York, Chichester, 2000.
10. van de Waterbeemd, H.; Smith, D. A.; Beaumont, K.; Walker, D. K. *J. Med. Chem.* **2001**, *44*, 1314, and references therein.
11. Springthorpe, B.; Strandlund, G. World Patent WO 2005/037052.
12. Jamieson, C.; Moir, E. M.; Rankovic, M.; Wishart, G. *J. Med. Chem.* **2006**, *49*, 5029.
13. (a) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeny, P. J. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3; (b) Wenlock, M. C.; Austin, R. P.; Barton, P.; Davis, A. M.; Leeson, P. D. *J. Med. Chem.* **2003**, *46*, 1250; (c) Leeson, P. D.; Davis, A. M. *J. Med. Chem.* **2004**, *47*, 6338.