Bioorganic & Medicinal Chemistry Letters 19 (2009) 2844–2851

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Defining optimum lipophilicity and molecular weight ranges for drug candidates—Molecular weight dependent lower log*D* limits based on permeability

Michael J. Waring *

AstraZeneca, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, United Kingdom

ARTICLE INFO

Article history:
Received 3 February 2009
Revised 20 March 2009
Accepted 21 March 2009
Available online 27 March 2009

Keywords:
Permeability,
Lipophilicity
Molecular weight
Caco-2

ABSTRACT

Analysis of a large, structurally diverse Caco-2 permeability dataset using a variety of statistical techniques suggests that log*D* and molecular weight are the most important factors in determining the permeability of drug candidates. The limit for log*D* is shown to be dependent on molecular weight. These limits are shown to be potentially superior to current guidelines in increasing the chances of finding highly permeable compounds. When combined with suggested upper limits for lipophilicity suggested in the literature based on the avoidance of toxicology and other adverse effects, this helps define a lipophilicity range that is optimum for drug candidates.

© 2009 Elsevier Ltd. All rights reserved.

It is well established that the lipophilicity of a molecule plays a crucial role in determining its suitability as a drug candidate. Seminal work by Lipinski[1] and others[2] in the late 90's went some way to establishing useful guideline regions to target in order to achieve acceptable oral exposure as part of the 'rule of 5'. For the most part, the emphasis has been on the avoidance of high lipophilicity (*c*log*P* above 5) due in part to the tendency of medicinal chemistry in general and combinatorial libraries in particular to produce compounds that are more lipophilic than optimal. More recently an increasing body of evidence is emerging that suggests that high lipophilicity increases the chances of compounds exhibiting adverse toxicological outcomes in addition to poor ADME and that these chances increase significantly, well below the *c*log*P* of 5 limit.[3–7]

In response to this, there is an increasing trend for medicinal chemistry to target areas of lower lipophilicity and, as a consequence, we would expect that the permeability of drug candidates might be compromised. It has been known for some time that permeability within a series diminishes with increasing polarity. This problem was mentioned by Lipinski[1] but was dismissed since it was not generally an issue for medicinal chemistry at the time. Nevertheless, permeability is accounted for in the Lipinski rules using other descriptors (donor count <5, acceptor count <10, molecular weight <500 Da).

A considerable number of other studies have been conducted in this area using in vivo fraction absorbed (F_{abs}), bioavailability and/or in vitro cell permeability data. In general, the dependence on molecular size, hydrogen bonding and lipophilicity have been established to a greater or lesser degree. Egan has developed a computational model that discriminates high, medium and low levels of absorption.[8] This model uses *A*log*P*98[9] and polar surface area (PSA) as key descriptors with well-absorbed compounds grouping in a region with *A*log*P*98 in the region of −1 to 5.9 and PSA less than 132 Å.[2] Veber has proposed a different set of criteria based on rat oral bioavailability data from the GlaxoSmithkline database which state that either PSA $\leqslant$ 140 Å^2 and rotatable bond count $\leqslant$10 or, alternatively, donor + acceptor count $\leqslant$ 12 and rotatable bond count $\leqslant$10 lead to increased chances of achieving a rat bioavailability of greater than 10%.[10] Of course, these rules do not deal directly with permeability but it is likely that permeability is a major factor in determining these limits since the link between these descriptors and permeability is much more established than it is for other bioavailability determining parameters such as first-pass hepatic metabolism. Indeed, Veber has shown that, whilst high in vitro cellular permeability (Caco-2 A to B P_{app} values >100 nm s^{-1}) do not guarantee high bioavailability, it is unlikely to be achieved for P_{app} values below this limit. Martin has proposed 'A Bioavailability Score' based on the analysis of rat oral bioavailability and Caco-2 permeability of compounds in the Abbott database.[11] These are more complex rules based on charge (at pH 6), PSA and Lipinski's rules. In summary, they predict that high perme-

* Tel.: +44 0 1625 230942; fax: +44 0 1625 516667.
E-mail address: mike.j.waring@astrazeneca.com

0960-894X/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2009.03.109

ability will be more likely for non-acids when Lipinski's rules are obeyed but for acids, permeability is better discriminated by PSA. From this, a score is derived which represents a probability of a compound having >10% rat bioavailability or measurably high Caco-2 permeability.

One potential drawback of these approaches in predicting permeability is that they either neglect, or at best account poorly for, ionisation. In general it is accepted that in the majority of cases the permeability neutral form of an ionisable compound is the most significant contributor to absorption[12] although studies by van de Waterbeemd have established that the ionised forms of some compounds are also capable of crossing membranes.[13,14]

Overall permeability, both in vitro and in vivo can be considered to be the sum of passive (diffusion driven) and active (transporter mediated) processes. The latter can effect both influx and efflux. In particular P-glycoprotein (PGP) mediated efflux is widely known to have a significant effect on absorption potential. Considering active efflux alone, Petrauskas has proposed the 'rule of four',[15] which states that compounds are likely to be efflux substrates if they have a hydrogen bond acceptor count (sum of N and O atoms) $\geqslant 8$, molecular weight >400 and an acid with a $pK_a > 4$. Conversely, compounds are likely to be non-substrates if they have an acceptor count $\leqslant 4$, molecular weight <400 and a base with a $pK_a < 8$. Of course this leaves a significant proportion of drug-like compounds unclassified. It is interesting to note, however, that these parameters are closely related to those described for determining overall permeability. Lipophilicity is implicated indirectly in this study since compounds with high numbers of hydrogen bond acceptors are likely to have low lipophilicity.

In order to examine lower limits of lipophilicity, it is better to use in vitro permeability data since the use of in vivo data is compounded by the multi-parameter nature of the measurement. For instance, low F_{abs} could be a consequence of poor solubility, sub-optimal formulation or pre-systemic metabolic events as well as poor permeability. These compounding factors would have no, or at best different, relationships with the descriptors in question. The link between cell monolayer permeability and in vivo F_{abs} is well established.[16,17] By analysing Caco-2 data, both the passive and active components of permeability are accounted for since Caco-2 cells express a range of relevant efflux transporters including PGP.[18] Hence conclusion from these data will suggest drivers for overall permeability but the distinction between passive and active components will be difficult to elucidate. This should be acceptable since the total permeability of compounds is usually the primary concern for optimisation and provided this remains high, it can be assumed that the passive permeability is high and active efflux is minimised.

Data were extracted from AstraZeneca's in house database of pooled Caco-2 measurements. At the time of extraction, this consisted of measurements of 9571 individual compounds. Where multiple measurements for single compounds exist, these were reduced to single values using the geometric mean. Predominantly, these are compounds screened as part of drug discovery projects across all the AstraZeneca sites but run in a consistent protocol comparable to standard methods in the literature.[14] It has been shown[8] that it is not reasonable to pool data from assays run under different protocols since significant differences exist between the values obtained. However, these data, whilst tested at different sites by different operators, are tested under a consistent protocol. The values obtained span the full dynamic range of the assay (Fig. 1).

Relevant molecular descriptors *AZ*log*D*,[19] *c*log*P*,[20] PSA, hydrogen bond donor and acceptor counts (Lipinski method), rotatable bond count and molecular weight were generated for all the compounds in the dataset using AstraZeneca's in house 'c-lab' platform. The *AZ*log*D* algorithm is an in house protocol for calculating log*D* and, in our experience, is the most accurate method of calculation for drug candidates. The commercial ACD log*D* algorithm[21] predicts the measured log*D* (where available) for this dataset with less accuracy (see Supplementary data). However, since there is little systematic bias in either prediction, the ACD algorithm, or other more accessible methods could be used in the same manner provided they can be shown to predict the measured log*D* values of the compounds in question with reasonable accuracy. All statistical analyses were carried out using JMP (version 6.0).[22]

The 9571 compounds in the dataset span more than four orders of magnitude of P_{app} values (Fig. 1a). Due to the dynamic ranges of the assays some of these measurements ($n = 1229$) have values outside the dynamic range of the assay in which they were measured, these compounds are excluded from analyses on continuous data such as calculations of the mean. They are left in analysis of categorical data and for calculation of the median. The data are clearly not normally distributed with a higher proportion of the compounds having high permeability. This is most likely because there is a limit to the maximum permeation rate that can be achieved in these assays (postulated to be limited by diffusion through the stagnant water layer).[13,23,24] For this reason the data are probably best treated in a categorical rather than a continuous manner. The biopharmaceutics classification system (BCS)[25,26] states that for compounds to be classified as having high permeability they are required to have a >90% F_{abs} in vivo based on compounds with known human F_{abs} data, this corresponds to P_{app} values $\geqslant 100$ nm s^{-1} in assays such as this. It is worth emphasizing that values below this do not preclude high F_{abs} but do introduce increased risk that the F_{abs} when measured might be low and variable across species or across studies and crucially, when assessed preclinically, represent a risk that the human F_{abs} will be low. For this reason it would seem that this is a suitable value to aspire to and hence a reasonable point to partition the data. This is consistent with the studies by Veber[10] and also provides a reasonably even split with 58% of the measurements below this value and 42% above (Fig. 1b). For statistical analyses of the type described here, it is preferable to have roughly even populations of the two categories.

Each descriptor was plotted against every other descriptor independently and a correlation coefficient (r value) for each was determined (see Supplementary data). Reasonably strong correlations exist between a number of these pairs of descriptors. Care should be applied in subsequent analysis in determining if one of these correlated terms is more important than another and in the combination of such correlated variables in multivariate analysis. However, there are a number of combinations of descriptors that are not correlated to any significant degree and, crucially for subsequent findings, the correlation between *AZ*log*D* and molecular weight is very weak ($r = 0.1027$).

Each of these descriptors was plotted independently in a linear manner against the logP_{app} value (Fig. 2). The relationship between permeability and lipophilicity has been reported to be linear,[27] hyperbolic,[28–30] sigmoidal,[23,24,31,32] parabolic[33] and bilinear.[27,34] It would therefore be unreasonable to assume that any of these other descriptors would show a simple, linear relationship. Moreover, for large, structurally diverse datasets such as this, significant spread in the data should be anticipated due to the influence of substructure and varying functionality in the compounds in question. Consequently, the linear plots reveal very weak correlations between the logP_{app} and each of the descriptors. However, the slopes of the lines indicate that the underlying trends in the data are those that would be expected. Permeability is likely to be compromised at low log*D*/log*P*, high PSA, high donor count, high acceptor count, high rotatable bond count and high molecular weight.

Since it is unreasonable to expect such simple descriptors to predict the observed P_{app} value in any quantitative way, interpre-

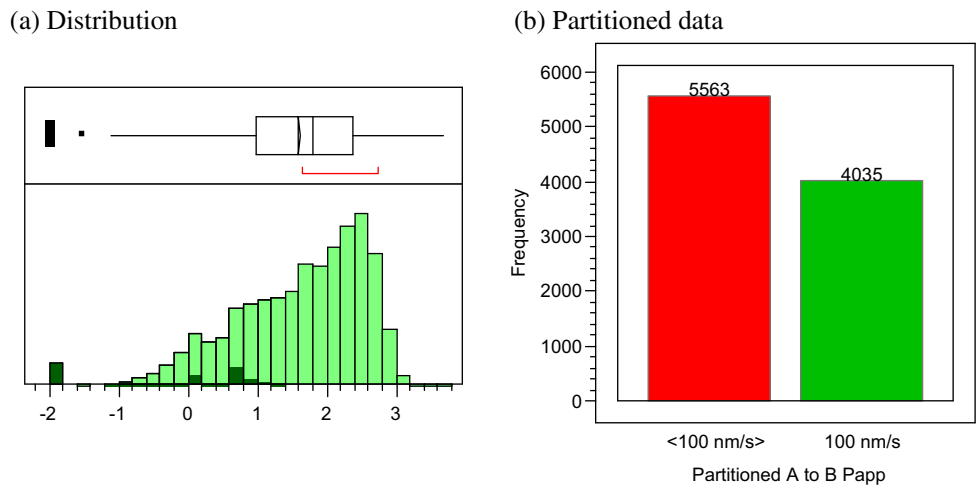

Figure 1. (a) Distribution of log*A* to *B* (P_{app}/nm s^{-1}) in the Caco-2 dataset, measurements outside the dynamic range of the assay are shaded and are removed for the purpose of calculating the mean and standard deviation but not the median. Mean = 0.677 (SD 0.871, *n* = 8865), median = 0.804 (*n* = 9571). (b) Dataset partitioned at P_{app} = 100 nm s^{-1}, 5563 (58%) have values <100 nm s^{-1}, *n* = 9598.

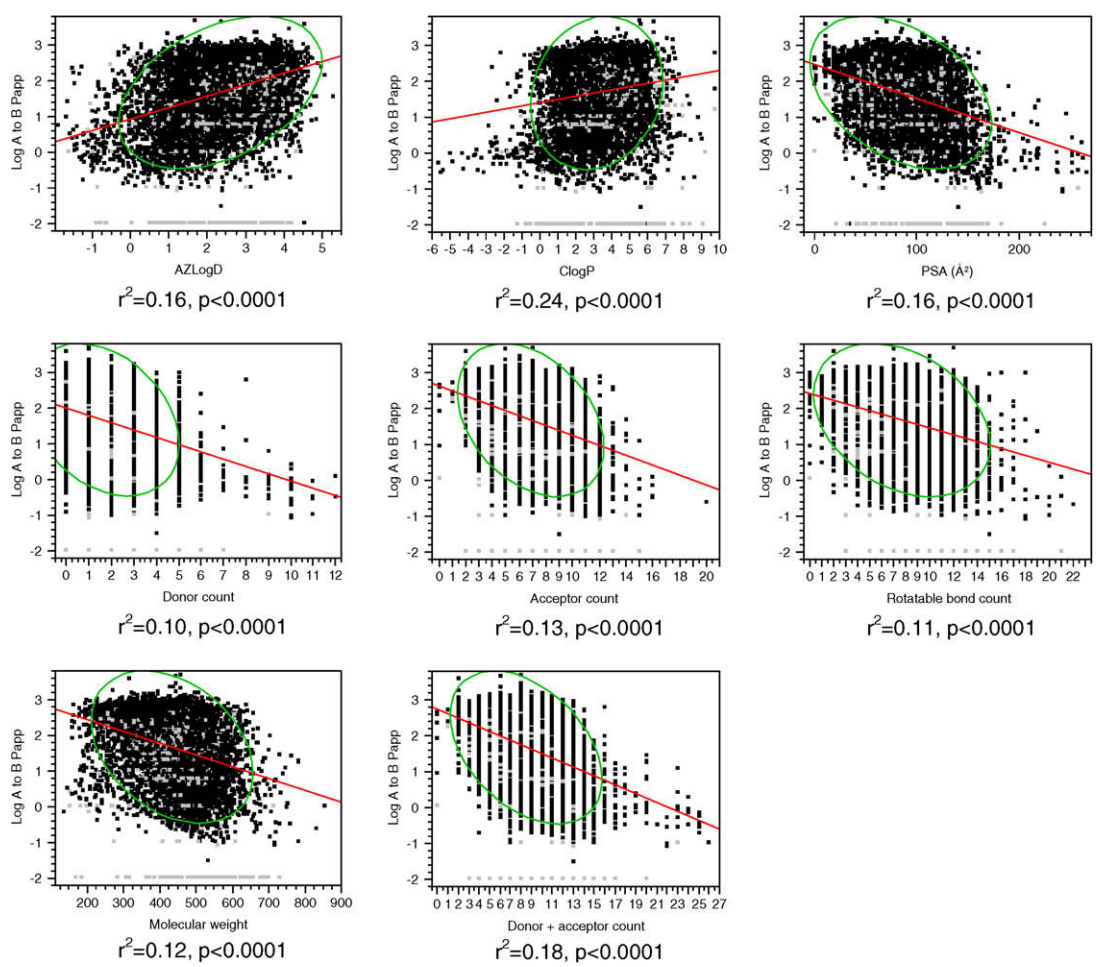

Figure 2. Linear relationships between 8 descriptors and log*A* to *B* P_{app}, *n* = 8865 (out of range data are shown in grey but excluded from the analysis). The legends show the correlation coefficients (r^2 values) and tests for significance (*p* values).

tation of these observations should be based on probability. Hence, a better way to examine these data is to carry out logistic regression[35] on the partitioned data. For such a large, structurally diverse dataset, the *y*-value of the line on the logistic plot can be interpreted as a probability of a compound with a given descriptor value achieving a P_{app} value above 100 nm s^{-1}.

Analysing the data in this way (Fig. 3) reveals a more obvious dependence of P_{app} on these descriptors. It can now be seen unambiguously that permeability is compromised at low *AZ*log*D*, high PSA, high donor count, high acceptor count, high rotatable bond count and high molecular weight. There is no dependence on *clog P*, suggesting as expected[13,14] that log*D*, rather than log*P*, is the

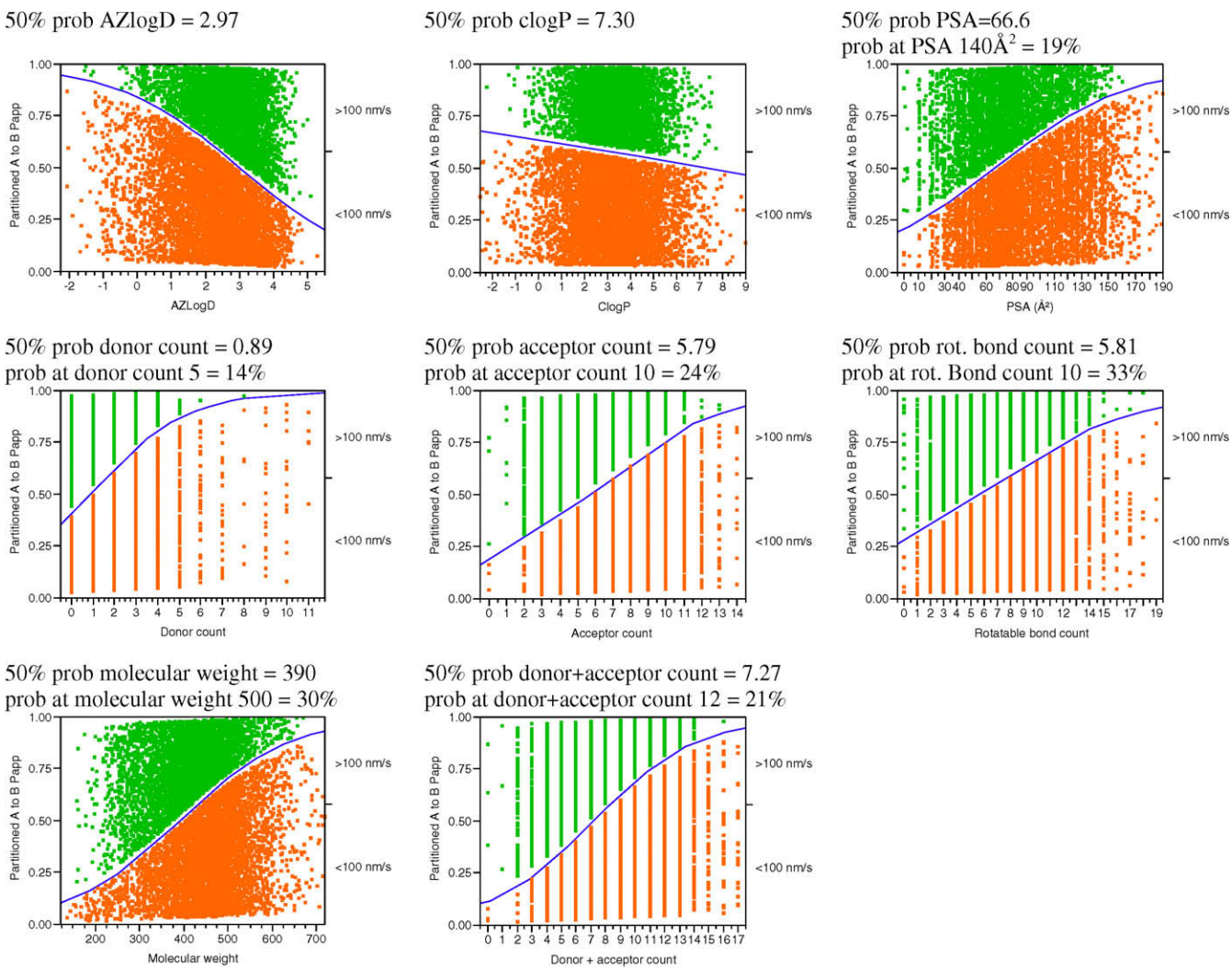

Figure 3. Logistic plots versus each single descriptor. Compounds with A to B P_{app} values >100 nm s^{-1} are shown in green, those <100 nm s^{-1} in red. The legend shows the value at which 50% probability occurs and, where appropriate, the probability of achieving A to B P_{app} values >100 nm s^{-1} at the Lipinski and Veber limits. The points are jittered randomly in the *y*-direction to indicate the relative populations, $n = 9598$.

important determinant of permeability. Moreover, this implies that the use of log*D* in place of log*P* might be sufficient to account for the ionisation of the compounds and there should be no need to examine the different ionisation classes separately. In actual fact, it may not be appropriate to categorise based on ionisation since the differing pK_a values within a class will affect the P_{app} differently. Using log*D* accounts for this since pK_a in conjunction with log*P* should affect the log*D* in the same way as the P_{app} provided the permeability of the ionised compound is negligible compared to that of its neutral form.

In addition, this analysis shows that the chances of achieving high permeability at the Lipinski limits for molecular weight and donor and acceptor counts as well as those for PSA, (donor + acceptor) count and rotatable bond count proposed by Veber are all low; between 14% and 33% in all cases, emphasising the benefit of operating well within, rather than close to, these limits.

Another observation from the logistic plots is that the probabilities change in a reasonably shallow fashion over the dynamic descriptor ranges. Hence application of hard rules as to optimum space should be treated with caution. However, such rules are appealing due to their simplicity and, provided these caveats are accepted, can serve as useful guidelines for compound design. One potential way to help define such limits is to use recursive partitioning. In this case, a computational algorithm is used to determine where the most appropriate point to cut the continuous data lies. By including all the above descriptors in such a process, it is also possible to determine objectively which of all the descriptors is most discriminating statistically.

Analysing the partitioned data using recursive partitioning including molecular weight, *AZ*log*D*, *c*log*P*, PSA, donor count, acceptor count and rotatable bond count as parameters shows that molecular weight and *AZ*log*D* are found to be the most discriminating factors (Fig. 4). The first split in the data occurs at molecular weight = 414 with 62% of compounds having a molecular weight below this value being poorly permeable. The second split discriminates on *AZ*log*D*, requiring a value of >1.3 for high permeability in the low molecular weight class and a value of >2.4 for the high molecular weight class. Further splitting of the data continues, at least initially, to subdivide on molecular weight and *AZ*log*D* rather than introducing new parameters.

This analysis suggests that molecular weight and *AZ*log*D* are the most important contributors, of those under consideration, to permeability. There is a 74% chance of achieving high permeability for compounds with molecular weight <414 and *AZ*log*D* > 1.3.

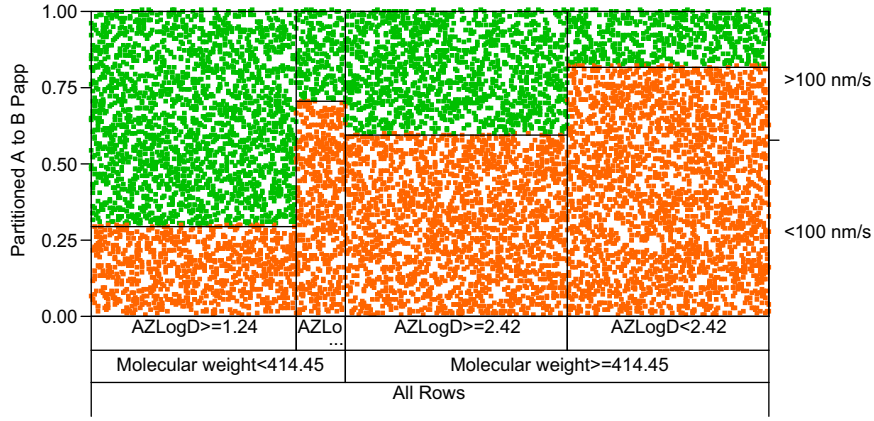

Figure 4. Recursive partitioning analysis of the partitioned P_{app} data. Compounds with A to B P_{app} values >100 nm s^{-1} are shown in green, those <100 nm s^{-1} in red, $n = 9598$.

Conversely, for a compound with molecular weight >414 and *AZ*log*D* < 2.4, there is an 81% chance of low permeability. It is also apparent from this that the influences of molecular weight and log*D* are not independent and that the log*D* threshold for high permeability occurs at a lower value for compounds of lower molecular weight. Such phenomena have been reported previously for different datasets.[11,14]

In order to investigate this further, the data were subdivided into even molecular weight bands, from 300 to 500 in increments of 50, and the logistic regressions with *AZ*log*D* were repeated for each molecular weight subset (Fig. 5). It is clear from this that the lipophilicity limit changes significantly over this molecular weight range. Moreover, perhaps the greatest changes occur precisely in the range that drug discovery projects usually operate.

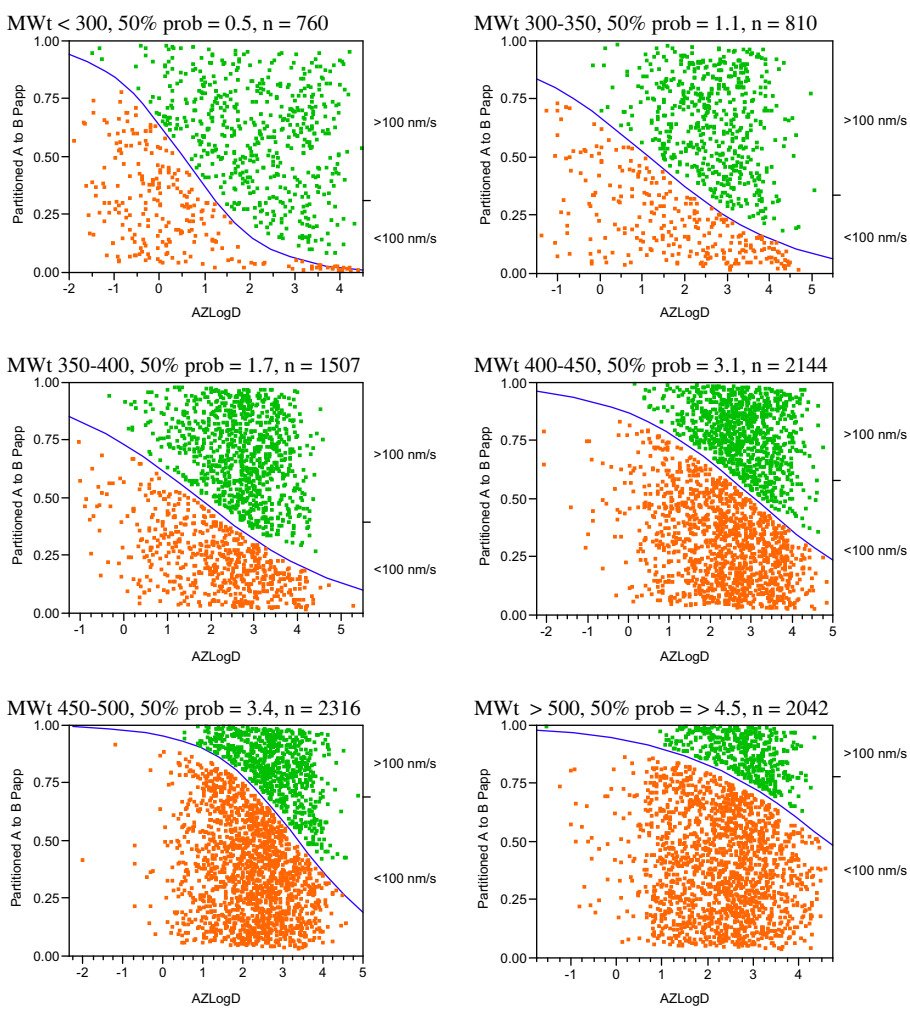

Figure 5. Logistic plots versus *AZ*log*D* for individual molecular weight bands.

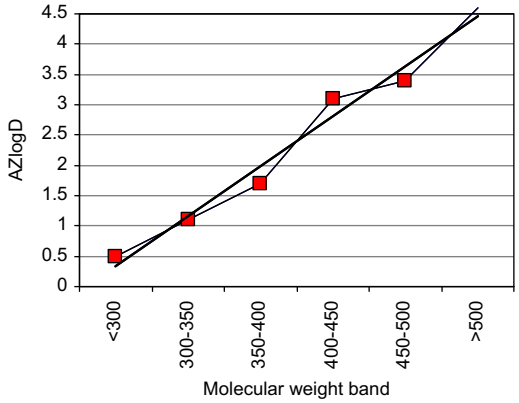

Figure 6. The required *AZ*log*D* to get a 50% chance of high permeability for a given molecular weight band. The bold line indicates the best linear fit.

For instance the 50% probability for the 350–400 molecular weight category is occurs at *AZ*log*D* = 1.7. For the next group (400–450) this occurs at a significantly higher value (*AZ*log*D* = 3.1).

It is possible that the apparent greater increase in this range is simply noise in the data; a graphical representation of how the required log*D* increases by molecular weight band (Fig. 6) reveals a roughly linear increase over the range studied. However, it is noticeable that this greatest change occurs at the molecular weight limit reported to be relevant for increased active efflux liability. This would be consistent with compounds requiring higher lipophilicity to achieve higher passive permeability to overcome active processes in addition to the detrimental effect of higher molecular weight on the passive component. Indeed, analysis of parallel artificial membrane permeability assay (PAMPA)[36] data in a similar manner reveals that the log*D* limit also increases with molecular weight band although somewhat less markedly overall (Table 1). The limit remains constant in the relevant (300–500) molecular weight range suggesting that active efflux may have a significant influence in this range.

From these plots, it is possible to determine guidelines based the 50% probability values that could be used as 'rules' in the same way as those proposed by Lipinski,[1] Veber[10] and Martin[11] (Table 2). These derived guidelines classify our permeability data with a reasonable level of success (Fig. 7a). Of the compounds predicted to be poor by these rules, 72% (4474/6196) had P_{app} values <100 nm s^{-1} in the Caco-2 assay. Of those predicted to be good, 68% (2305/3383) had P_{app} values $\geqslant$100 nm s^{-1}. Since this dataset was extracted, a further 401 compounds have been tested. These compounds serve as a useful independent test set for these rules. For this dataset, the rules discriminate similarly well with 82% (107/130) of those passing the criteria having P_{app} values $\geqslant$100 nm s^{-1} and 70% of those failing having P_{app} values <100 nm s^{-1} (Fig. 7b).

To compare the ability of these criteria to discriminate, the three sets of literature criteria (Lipinski, Veber and Martin) were also applied to this dataset (Fig. 8). In this case failure of any one of the Lipinski rules was treated as an overall failure, for the Martin criteria, a Bioavailability Score of 0.11 or 0.17 was treated as a fail and a score of 0.55, 0.56 or 0.85 was treated as a pass. The rules derived here identify the poorly permeable compounds with a similar level of success to the Veber, Martin and Lipinski rules (Table 3, 70–80% in all cases) but perform significantly better in selecting the highly permeable compounds. Of the compounds meeting the criteria, 68% are actually highly permeable, compare this to 46–48% for the Veber, Martin and Lipinski rules. Hence, application of these rules should lead to significantly increased chances of success in identifying highly permeable compounds with a similar proportion of false negatives and should represent an advance on current design criteria.

The application of the Petrauskas 'rule of four' for efflux[15] shows some discrimination in the overall P_{app} values. Of the compounds classified as being likely efflux substrates, 78% have P_{app} values <100 nm s^{-1} and 88% of the likely non-substrates have P_{app} values >100 nm s^{-1} although very few compounds actually fall into this latter category (Fig. 9). Large proportion of the data (57%) are not classified by these rules and remain undetermined and these indeterminate compounds show a roughly equal split between high and low P_{app} categories. As stated previously, it is difficult to uncouple the active and passive components of permeability and the descriptors proposed for both are clearly very similar but the log*D* and molecular weight criteria proposed perform similarly to the rule of four in predicting high overall permeability and have the added advantage that they classify all compounds.

The findings presented here should assist with the design of compounds with improved permeability. Molecular weight and lipophilicity (as measured by log*D* rather than log*P*) appear to be at least comparable with, and probably superior to, any other simple molecular descriptors in assessing permeability. This is convenient given the contribution that log*D* and molecular weight have been shown to make to other ADMET properties and to overall drug-likeness. In addition, these two factors are easier to factor into compound design than some of the more complex parameters. In our experience, the problem of multi-parameter optimisation of drug molecules becomes simpler the more parameters can be linked to mutual descriptors.

What is striking about this is the quantification of the log*D* limits compared to the upper limits proposed based on toxicological outcomes. Leeson and Springthorpe have proposed that in order to avoid high levels of promiscuity *c*log*P* should be kept below 3.[4] The same limit has been proposed based on observations made by Price et al. at Pfizer based on adverse in vivo toxicological outcomes.[5] Gleeson has proposed that *c*log*P* should be kept below 4 and molecular weight below 400 based on an analysis of a wide range of secondary screening data.[7] Of course, all of these studies deal with log*P* rather than log*D* values but these studies also imply upper log*D* limits, albeit different depending on the degree of ionisation.

If log*P* is the important factor, it might be concluded that the optimal lipophilicity space for ionisable compounds is small and

Table 1
*AZ*log*D* values required to obtain a 50% probability of P_{app} values >100 nm s^{-1} in the parallel artificial membrane permeability assay (PAMPA) (n = 1922)

Molecular weight	*AZ*log*D*
<300	1.7 (n = 140)
300–350	2.2 (n = 150)
350–400	2.6 (n = 333)
400–450	2.7 (n = 420)
450–500	2.5 (n = 517)
>500	>4.5 (n = 362)

Table 2
Permeability rules defining *AZ*log*D* limits required to achieve >50% chance of high permeability for a given molecular weight band

Molecular weight	*AZ*log*D*
<300	>0.5
300–350	>1.1
350–400	>1.7
400–450	>3.1
450–500	>3.4
>500	>4.5

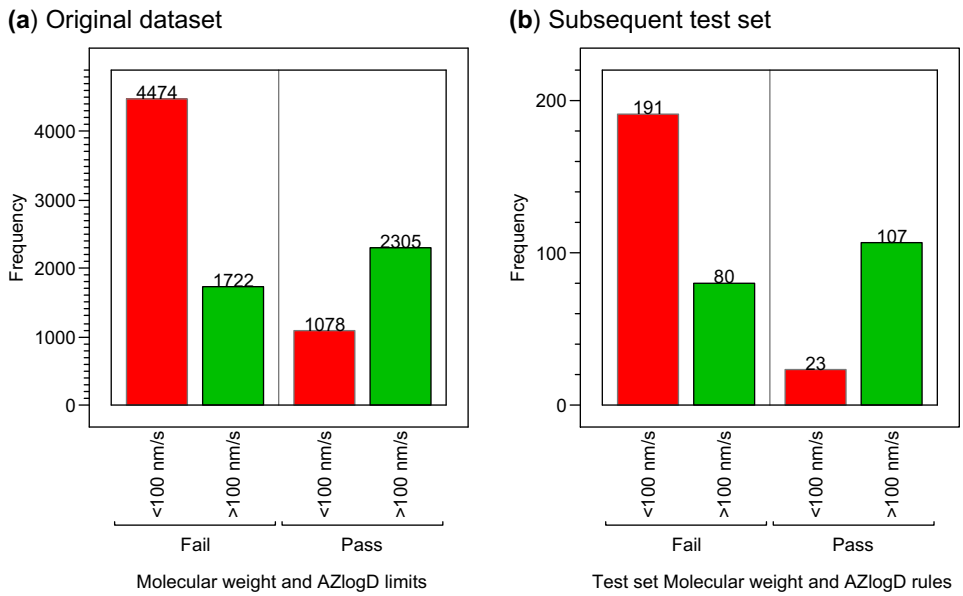

Figure 7. P_{app} categories of the compounds that pass and fail the *AZ*log*D* and molecular weight criteria showing the populations of those correctly and incorrectly classified for the original dataset (a) and for the subsequent test set (b). Compounds with measured A to B P_{app} values >100 nm s^{-1} are shown in green, those <100 nm s^{-1} in red.

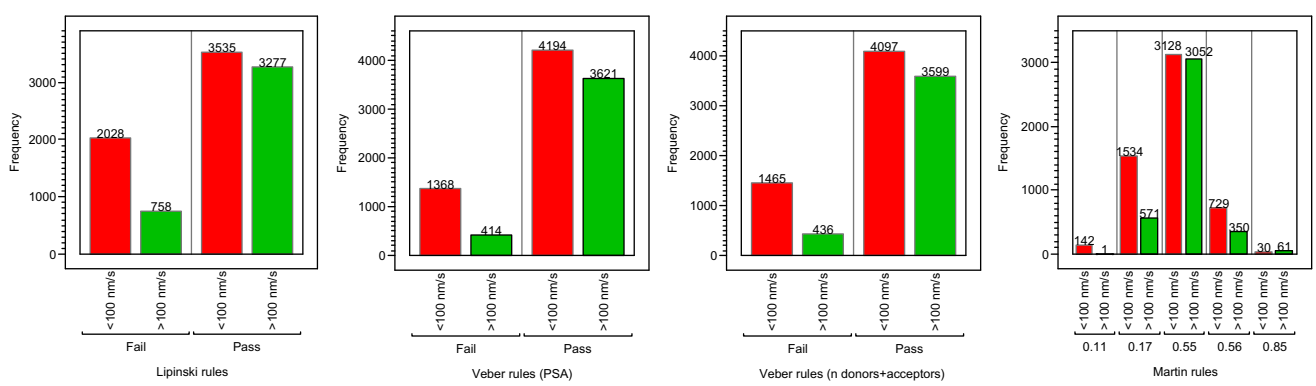

Figure 8. P_{app} categories of the compounds that pass and fail the Lipinski, Veber and Martin criteria showing the populations of those correctly and incorrectly classified for the original dataset. Compounds with measured A to B P_{app} values >100 nm s^{-1} are shown in green, those <100 nm s^{-1} in red.

Table 3
Summary of the percentages of compounds correctly classified by the molecular weight and *AZ*log*D* criteria compared to those of Lipinski, Veber and Martin

Rules	Status	Correct (%)
Molecular weight and *AZ*log*D*	Fail	72
	Pass	68
Lipinski	Fail	73
	Pass	48
Veber (PSA)	Fail	77
	Pass	46
Veber (donor + acceptor count)	Fail	77
	Pass	47
Martin	Fail (0.11 and 0.17 probability)	75
	Pass (0.55, 0.56 and 0.85 probability)	47

potentially non-existent—a basic compound with a pK_a of 8.4 would require a log*D* value of 2 to have a log*P* of 3 and these findings would suggest that for molecular weights above 450, significant permeability attrition would be anticipated.

Of course, all these conclusions are based on probability analysis, it is clear that it is possible to find compounds with log*P* values >3 that do not show adverse toxicology and there are significant numbers of compounds that are known to be sufficiently permeable outside the limits defined here. In addition, the impact of active uptake in vivo is known to allow for absorption of very polar compounds and targeting these mechanisms may allow a means of circumventing these constraints.

The majority of studies of this type have tended to focus thus far on upper limits of lipophilicity. This study establishes that a lower limit also exists. The issues associated with high lipophilicity are manifold.[1,2,4–7] The issues with low lipophilicity are likely to be fewer, with permeability being the most obvious, but no less important. High quality drug candidates need to fulfill all of these criteria and the resolution of those that are contradictory are key aspects of optimisation. This study suggests that optimum lipophilicity space is better considered as a range with upper and lower limits rather than a simple upper limit. Combining this with the literature studies proposing upper lipophilicity limits indicates that this range is very narrow. The combined impact of molecular

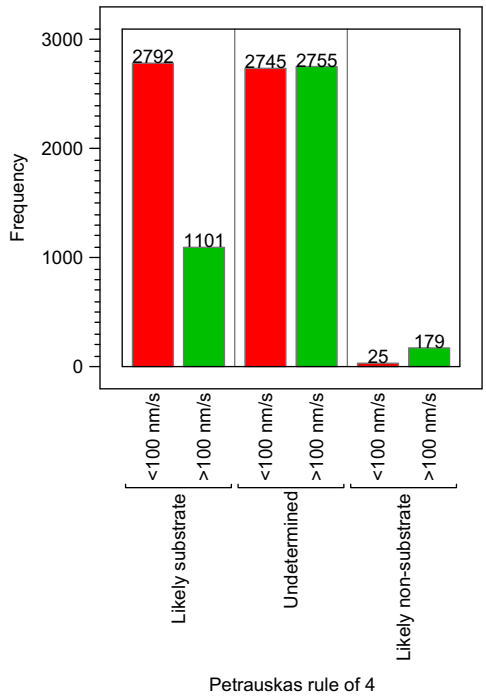

Figure 9. Performance of the Petrauskas 'rule of four' in predicting overall P_{app} values. Compounds with measured A to B P_{app} values >100 nm s^{-1} are shown in green, those <100 nm s^{-1} in red.

weight and log*D* imply that control of molecular weight is a means of increasing the width of this optimum range.

Acknowledgements

The author thanks Drs. A. L. Gill, C. Johnstone and D. McKerrecher for helpful discussions relating to this work.

Supplementary data

Supplementary data (plots of the *AZ*log*D* and ACD log*D* values versus measured data, a table showing the cross-correlations of the descriptors and a representative subset of the data containing all relevant parameters are available) associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2009.03.109.

References and notes

1. Lipnski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **1997**, *23*, 3.
2. Navia, M. A.; Chaturverdi, P. R. *Drug Discovery Today* **1996**, *1*, 179–189.
3. Wenlock, M. C.; Austin, R. P.; Barton, P.; Davis, A. M.; Leeson, P. D. *J. Med. Chem.* **2003**, *46*, 1250–1256.
4. Leeson, P. D.; Springthorpe, B. *Nat. Rev. Drug Disc.* **2007**, *6*, 881–890.
5. Hughes, J. D.; Blagg, J.; Price, D. A.; Bailey, S.; DeCrescenzo, G. A.; Devraj, R. V.; Ellsworth, E.; Fobian, Y. M.; Gibbs, M. E.; Gilles, R. W.; Greene, N.; Huang, E.; Krieger-Burke, T.; Loesel, J.; Wager, T.; Whiteley, L.; Zhang, Y. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4872–4875.
6. Waring, M. J.; Johnstone, C. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1759–1764.
7. Gleeson, M. P. *J. Med. Chem.* **2008**, *51*, 817–834.
8. Egan, W. J.; Merz, K. M.; Baldwin, J. J. *J. Med. Chem.* **2000**, *43*, 3867–3877.
9. Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. *J. Phys. Chem. A* **1998**, *102*, 3762–3772.
10. Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. *J. Med. Chem.* **2002**, *45*, 2615–2623.
11. Martin, Y. C. *J. Med. Chem.* **2005**, *48*, 3164–3170.
12. Yoshida, F.; Topliss, J. G. *J. Med. Chem.* **2000**, *43*, 2575–2585.
13. Camenisch, G.; Folkers, G.; van de Waterbeemd, H. *Eur. J. Pharm. Sci.* **1998**, *6*, 321–329.
14. Camenisch, G.; Alsenz, J.; van de Waterbeemd, H.; Folkers, G. *Eur. J. Pharm. Sci.* **1998**, *6*, 313–319.
15. Didziapetris, R.; Japertas, P.; Avdeef, A.; Petrauskas, A. *J. Drug Targeting* **2003**, *11*, 391–406.
16. Artursson, P.; Karlsson, J. *Biochem. Biophys. Res. Commun.* **1991**, *175*, 880–885.
17. Mandagere, A. K.; Thompson, T. N.; Hwang, K. K. *J. Med. Chem.* **2002**, *45*, 304–311.
18. Maubon, N.; Le Vee, M.; Fossati, L.; Audry, M.; Le Ferrec, E.; Bolze, S.; Fardel, O. *Fund. Clin. Pharmacol.* **2007**, *21*, 659–663.
19. Bruneau, P.; McElroy, N. R. *J. Chem. Inf. Model.* **2006**, *46*, 1379–1387.
20. http://www.daylight.com.
21. SAS Institute. JMP Statistical Software. http://www.jmp.com/software.
22. http://www.acdlabs.com.
23. Ho, N. F. H.; Park, J. Y.; Morozowich, W.; Higuchi, W. I. In *Design of Bioparmaceutical Properties through Prodrugs and Analogs*; Roche, E. B., Ed.; APhA/APS: Washington, DC, 1977; pp 136–227.
24. Ho, N. F. H.; Park, J. Y.; Ni, P. F.; Higuchi, W. I. In *Animal Models for Oral Drug Delivery in Man: In Situ and In Vivo Approaches*; Crouthamel, W. G., Sarapu, A., Eds.; APhA/APS: Washington, DC, 1983; pp 27–106.
25. http://www.fda.gov/cder/OPS/BCS_guidance.htm
26. Martin, A.; Bustamante, P.; Chun, A. *Physical Pharmacy: Physical Chemical Principles in the Pharmaceutical Sciences*; Lea and Febiger: Philadelphia, 1993.
27. Kakemi, K.; Arita, T.; Hori, R.; Konishi, R. *Chem. Pharm. Bull.* **1967**, *17*, 1534–1539.
28. Wagner, J. G.; Sedman, A. J. *J. Pharmacokinet. Biopharm.* **1973**, *1*, 23–50.
29. Garringues, T. M.; Pérez-Varona, A. T.; Climent, E.; Bermejo, M. V.; martin-Villdore, A.; Plá-Delfina, J. M. *Int. J. Pharm.* **1990**, *64*, 127–138.
30. Lee, A. J.; King, J. R.; Barrett, D. A. *J. Control. Release* **1997**, *45*, 141–151.
31. Adson, A.; Burton, P. S.; Raub, Th. J.; Barshun, C. L.; Audus, K. L.; Ho, N. F. H. *J. Pharm. Sci* **1995**, *84*, 1197–1204.
32. Camenisch, G.; Alsenz, J.; van de Waterbeemd, H.; Folkers, G. *Eur. J. Pharm. Sci.* **1998**, *6*, 313–319.
33. Hansch, C. *Acc. Chem. Res.* **1969**, *2*, 232–239.
34. Kubinyi, H. *Arzneim. Forsch. (Drug Res.)* **1976**, *26*, 1991–1997.
35. Hosmer, D. W.; Lemeshow, S. *Applied Logistic Regression*, 2nd Ed.; Wiley: New York, Chichester, 2000.
36. Kansy, M.; Senner, F.; Gurbenator, K. *J. Med. Chem.* **1998**, *41*, 1007–1010.