Computational methods for the prediction of 'drug-likeness'

David E. Clark and Stephen D. Pickett

Recently, one of the key trends in the pharmaceutical industry has been the integration of what have traditionally been considered 'development' activities into the early phases of drug discovery. The aim of this paradigm shift is the prompt identification and elimination of candidate molecules that are unlikely to survive later stages of discovery and development. In this review, the authors examine the growing role that is being played by computational methods in this filtering process, with a particular focus on the prediction of intestinal absorption and blood–brain barrier penetration.

Much has been written recently concerning the impact of combinatorial chemistry and high-throughput screening (HTS) on drug discovery. However, there is a growing realization that, given the vast size of organic chemical space (possibly $>10^{18}$ compounds), drug discovery cannot be reduced to a simple 'synthesize-and-test' lottery. Researchers must somehow 'load the dice' to give the compounds they are making the best possible chance of becoming drugs. In other words, there is a need to identify and/or design a subset of 'drug-like' molecules from the vast expanse of what could possibly be synthesized.

To this end, a battery of *in vitro* ADME (absorption, distribution, metabolism, excretion) screens has been implemented in most pharmaceutical companies with the aim of discarding compounds in the discovery phase that are likely to fail further down the line[1]. The motto in the industry is now 'fail fast, fail cheap'. Thus, candidate molecules are checked both for their absorption properties using Caco-2 (a human intestinal epithelial cell line derived from a colorectal carcinoma)[2] or MDCK (Madin–Darby canine kidney)[3] cell monolayers and for their susceptibility to metabolic degradation using liver microsomes or hepatocytes[4].

As valuable as these experimental filters are, they do have some limitations. For example, they require physical samples of compounds for testing and, despite significant technical advances, they remain time-consuming and resource-intensive. Thus, there is currently much interest in the development and application of computational methods for predicting 'drug-likeness'[5]. Such methods could be applied to *virtual* compounds or libraries permitting the rapid and cost-effective elimination of poor candidates prior to synthesis.

The current state-of-the-art in terms of the prediction of drug-likeness will now be reviewed, with particular reference to computational methods for the prediction of:

- Drug-likeness in a general sense
- Intestinal absorption
- Blood–brain barrier (BBB) penetration.

It should be noted that with regard to the latter two categories, only the prediction of absorption by passive mechanisms will be considered. Also, while of much interest and importance, the prediction of both active transport by various carrier systems and the metabolic fate of compounds will not be covered in this review.

Prediction of general drug-likeness

Initial strategies towards this goal involved the use of computational filters to remove compounds deemed to be chemically unsuitable for screening purposes[6–9]. Such filters typically incorporate substructure searches for toxic or reactive groups (such as acid halides) and can include limits on MW and the number of rotatable bonds. These

David E. Clark* and **Stephen D. Pickett**, Computer-Aided Drug Design, Rhône-Poulenc Rorer Ltd, Dagenham Research Centre, Rainham Road South, Dagenham, Essex, UK RM10 7XS. *tel: +44 020 8919 3353, fax: +44 020 8919 2029, e-mail: david-e.clark@rp-rorer.co.uk

 1359-6446/00/$ – see front matter PII: S1359-6446(99)01451-8

methods provide a very useful initial filtering of any data set and can be used prior to further data analysis. However, by their very nature, instead of addressing the question of what makes a drug, these filters attempt to eliminate definite non-drugs. A more sophisticated approach to predicting drug-likeness is to examine two sets of compounds, one containing known drugs and the other comprising compounds known (or presumed) not to be potential drugs. The properties or characteristics of the former set that differentiate them from the latter can then be elucidated using computer methods. This type of approach has already been applied by several research groups.

Genetic algorithm-based approaches

A genetic algorithm-based approach has been used[10] to calculate weights that describe the differential occurrences of simple, one-dimensional (1-D) molecular properties (MW, the numbers of hydrogen bond donors and acceptors, rotatable bonds and aromatic rings) and a topological shape descriptor in compounds from the World Drug Index (WDI, Derwent Information, http://www.derwent. com), assumed to be drug-like, and compounds from the SPRESI database (Daylight Information Systems, http://www.daylight.com), assumed to be non-drug-like. The resulting biological activity profiles are reasonably effective at discriminating between the two classes, especially considering the simple nature of the descriptors used to characterize the molecules. A method incorporating this work has been used at GlaxoWellcome (Stevenage, Hertfordshire, UK) to filter compounds prior to HTS (Ref. 9).

Neural network-based approaches

Researchers at Vertex (Cambridge, MA, USA) trained a Bayesian neural network to recognize drug-like molecules using the Comprehensive Medicinal Chemistry (CMC, MDL Information Systems, http://www.mdli.com) and Available Chemicals Directory (ACD, MDL Information Systems) as their sets of drugs and non-drugs, respectively[11]. Using a combination of simple 1-D descriptors and the ISIS (MDL Information Systems) 2-D substructural keys to characterize the molecules, the neural network classified 90% of the CMC compounds correctly as drugs while incorrectly predicting only 11% of the ACD to be drug-like. The same network also classified nearly 80% of another drug database used as a test set, the MDDR (MDL Drug Data Report, MDL Information Systems), as drug-like.

In similarly motivated, but independent, work carried out at BASF AG (Ludwigshafen, Germany), a feed-forward neural network was trained using molecules from the WDI as a drug-like set and the ACD as a non-drug-like set[12]. In this work, the compounds were characterized using atom-type descriptors originally developed for predicting octanol/water partition coefficients. The resulting scoring scheme succeeded in classifying 77% of the WDI correctly as drugs and 83% of the ACD correctly as non-drugs. The method is very fast, being able to classify approximately 100 compounds per second on a modern workstation. More recently, a different classification technique (recursive partitioning) has been applied to these data sets with similarly successful results [Wagener, M. and van Geerestein, V.J. (1999) Analysing large data sets with decision trees: Discriminating between potential drugs and nondrugs. *Abstracts of the 217th American Chemical Society Meeting*. 21–25 March, Anaheim, CA, USA].

Limitations

Thus, it would seem possible to develop rapid computational filters that can distinguish between compounds that are drug-like and non-drug-like. The lack of validated sets of drugs and non-drugs is a disadvantage for these methods and there are undoubtedly compounds that do not easily fall into either category. However, given the time and expense of conducting such a validation exercise, there is little choice but to use sets like those already mentioned. Another criticism is that the classifiers can only recognize those compounds that resemble existing drugs as drug-like – compounds from completely new classes could be misclassified and this could hinder innovation[13]. In terms of applications, these techniques will probably be most useful in lead discovery where they could be used for prioritizing compounds for HTS or for purchasing from external suppliers. They could also be applied for the design of biased combinatorial libraries. Their use in lead optimization is more restricted, as they can only give a yes/no answer to the question 'is this a drug-like molecule?' At this stage in the drug discovery process, it is necessary to resort to more specific approaches such as those described in the following sections.

Prediction of intestinal absorption

For most drugs, the preferred route of administration is by oral ingestion. Researchers have therefore sought to delineate the physicochemical properties that favour intestinal absorption[14,15] and develop computational methods for its prediction[16].

The 'rule-of-five'

Probably the best-known of these methods is the 'rule-of-five'[17] devised by Lipinski and coworkers at Pfizer

(Groton, NJ, USA) from an analysis of 2245 drugs from the WDI believed to have entered Phase II trials. As implemented in the Pfizer registration system, the rule-of-five generates an alert (indicating possible absorption problems) for compounds where any two of the following conditions are satisfied:

- Molecular weight >500
- Number of hydrogen-bond acceptors >10
- Number of hydrogen-bond donors >5
- Calculated logP >5.0 (if ClogP is used) or >4.15 (if MlogP is used).

In this rule, any oxygen (O) and nitrogen (N) atoms are defined as hydrogen-bond acceptors and N–H or O–H groups are considered as hydrogen-bond donors. LogP refers to the octanol/water partition coefficient of a compound and is used as a measure of lipophilicity. Many approaches for calculating logP have been developed including ClogP (Daylight Information Systems, Mission Viejo, CA, USA) and MlogP (developed by Moriguchi and coworkers)[18]. An analysis of the CMC database by Amgen (Thousand Oaks, CA, USA) researchers[19] has reinforced some of these findings, particularly concerning the preferred ranges of MW and logP.

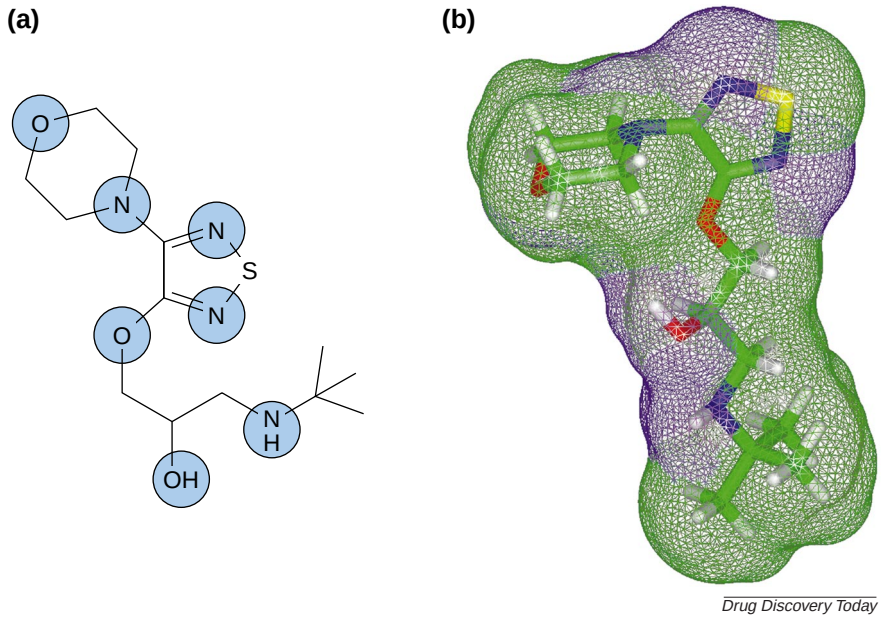

Figure 1. *The beta-blocker timolol, showing the atoms contributing to its polar surface area in (a) two dimensions (circled atoms) and in (b) three dimensions (molecular surface coloured purple).*

Predicting Caco-2 cell permeability

Instead of analyzing existing collections of drugs, another approach has been to model specific, absorption-related quantities. For example, the following QSPR (quantitative structure–property relationship) model was developed to describe the Caco-2 permeabilities of 17 diverse drug compounds[20]:

$$\log P_{app} = 0.008 \times MW - 0.043 \times PSA - 5.165 \quad (1)$$

Here, $\log P_{app}$ is the logarithm of the apparent permeability (cm s^{-1}) through the monolayer and the PSA indicates the polar surface areas of the compounds. The correlation coefficient (r) for this equation is 0.833. The equation indicates that permeability increases with increasing MW and decreasing PSA. This data set has also been studied using other computed descriptors and partial least squares (PLS) regression[21]. The results of this analysis also indicated the importance of minimizing properties associated with hydrogen bonding if good permeability is to be achieved. This conclusion is in accord with the observation that cellular permeability is primarily dependent on the desolvation potential of the polar functional groups in a molecule and only secondarily dependent on its lipophilicity[22].

The concept of using PSA as a predictor of absorption has recently attracted much interest. The PSA is defined as that part of the molecular surface that arises from oxygen or nitrogen atoms, or hydrogen atoms attached to nitrogen or oxygen atoms (see Fig. 1). Some workers also include sulphur (and hydrogen attached to sulphur) in their definitions. Early work by Palm and coworkers at Uppsala University[23] established a strong correlation ($r^2 = 0.99$) between the dynamic PSA (PSA_d) and the Caco-2 permeabilities of six beta-blockers (Fig. 2).

PSA_d is calculated by taking a Boltzmann-weighted average of the PSA values of all conformers within 2.5 kcal mol^{-1} of the lowest-energy conformer found during a conformational search (a relatively time-consuming procedure). A similar approach used by other workers showed that PSA_d correlated well with the Caco-2 permeability data of some beta-blockers and their ester prodrugs[24]. Lower

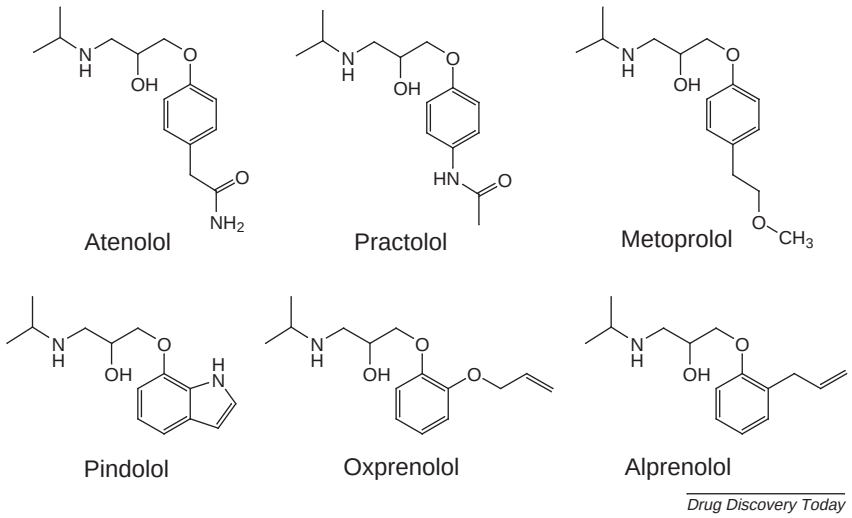

Figure 2. *The beta-blockers showing a good correlation between PSA and Caco-2 cell permeability[23].*

PSA_d was proffered as an explanation for the preferential permeabilities of the prodrugs. More recent work from the Uppsala group has examined the effect of computing PSA_d in different simulated environments (water, vacuum, chloroform) and concluded that, for most of the nine beta-blockers studied, the effect of altering the environment on PSA_d was small[25]. In addition, consideration of the non-PSA has been shown to be beneficial in modelling the Caco-2 permeabilities of an expanded set of 19 oligopeptides[26].

Predicting human fractional absorption

While Caco-2 monolayers can provide useful data for predicting intestinal absorption, they are only a model of the human absorption process. There is, therefore, much interest in generating predictive models based on (currently quite sparse) human absorption data. The initial work in this area, again by Palm and coworkers, involved a set of 20 drugs covering a wide range of fractional absorption values (%FA) in humans (Table 1)[27]. These compounds were carefully selected to ensure they were primarily absorbed by passive processes and to include a wide range of physicochemical properties. Using this set, a strong sigmoidal correlation between PSA_d and %FA ($r^2 = 0.94$) was demonstrated, suggesting that compounds with a PSA_d >140 $Å^2$ are likely to show poor absorption characteristics (%FA <10). Norinder and coworkers[28] have applied their MolSurf (Qemist AB, Karlskoga, Sweden) descriptors and PLS regression methods to this data set and arrived at similar conclusions to those reached with their Caco-2 study[21]. A novel molecular surface-based descriptor, known as the 'molecular hashkey', has also been correlated with these data using a backpropagation neural network[29].

An alternative approach using a genetic algorithm with a neural network scoring function and a larger training set of 76 compounds, resulted in the generation of a QSPR model relating %FA to several structure-derived descriptors[30]. The root-mean square errors (RMSE) predicted from this model were quite reasonable: 9.4% for the training set and 16.0% for a test set of ten compounds. However, the data set used in this work is heavily biased towards well-absorbed compounds (only 15 of the 86 compounds have %FA values of less than 50), reducing the capacity of the model to predict %FA values for less well-absorbed compounds. Furthermore, the data set contains compounds that are absorbed by active transport processes as well as passively absorbed compounds. This is problematic because it would seem unrealistic to expect to find a single model that will accurately predict such different physical processes. Passive absorption is dependent on the rate of diffusion of a compound, either between (paracellular) or across (transcellular) the cells comprising the epithelial barrier. This diffusion rate will, in turn, depend on the physicochemical properties

Table 1. The 20 drugs studied by Palm and coworkers, together with their mean fractional absorption data in humans[27]

Compound	[a]Mean %FA	Compound	Mean %FA
Metoprolol	102.0	Metolazone	64.0
Nordiazepam	99.0	Tranexamic acid	55.0
Diazepam	97.0	Atenolol	54.0
Oxprenolol	97.0	Sulpiride	36.0
Phenazone	97.0	Mannitol	26.0
Oxazepam	97.0	Foscarnet	17.0
Alprenolol	96.0	Sulfasalazine	12.0
Practolol	95.0	Olsalazine	2.3
Pindolol	92.0	Lactulose	0.6
Ciprofloxacin	69.0	Raffinose	0.3

[a] Fractional absorption.

of the compound. By contrast, each active transport mechanism requires more specific molecular recognition.

It has recently been shown how the time-consuming calculation of PSA_d can be avoided by using PSA values based on a single, rapidly calculated low-energy conformer[31]. On application of this method to the 20-compound data set shown in Table 1, an almost identical correlation was obtained ($r^2 = 0.94$; see Fig. 3). This faster computational protocol enables the rapid screening of large (virtual) data sets, such as combinatorial libraries. The criterion for poor absorption of PSA >140 Å^2 was robust when applied to the set of 76 compounds referred to earlier[30], if the actively transported compounds are excluded[31].

Predicting human effective permeability

A different measure of intestinal absorption in humans, effective permeability (P_{eff}), has also been investigated[32]. Thirteen passively absorbed compounds (Table 2) were considered in the study and their $\log P_{eff}$ values correlated with some calculated descriptors using PLS analysis. Notably, the simplest model with reasonable statistics ($q^2 = 0.82$, where q^2 is the cross-validated r^2; $r^2 = 0.85$) included a single-conformer PSA value together with a count of the number of hydrogen-bond donors (HBD):

$$\log P_{eff} = -2.546 - 0.011 \times PSA - 0.278 \times HBD \qquad (2)$$

As with Equation (1), this model indicates that decreasing the number of polar (hydrogen bonding) functional groups in a molecule favours passive absorption. The underlying physical explanation for this is that, in an aqueous environment (e.g. the intestine), the polar groups of a molecule will be solvated. Entry to the more lipophilic environment of the intestinal epithelium requires the desolvation of these groups. If this is too energetically expensive, then absorption will be hindered. It would seem from these results, that PSA is in some way describing the degree of solvation and, thus, the energy required for desolvation of a compound[33]. However, workers at Pharmacia and Upjohn (Kalamazoo, MI, USA) have recently failed to find a good correlation between PSA and the experimentally determined hydrogen-bonding potential for a set of peptides and peptidomimetics[34]. Thus, there is still much scope for further research in this area, perhaps looking at more sophisticated definitions of the PSA (Ref. 31) and accounting for the relative hydrogen-bond donor and acceptor strengths of molecules[35].

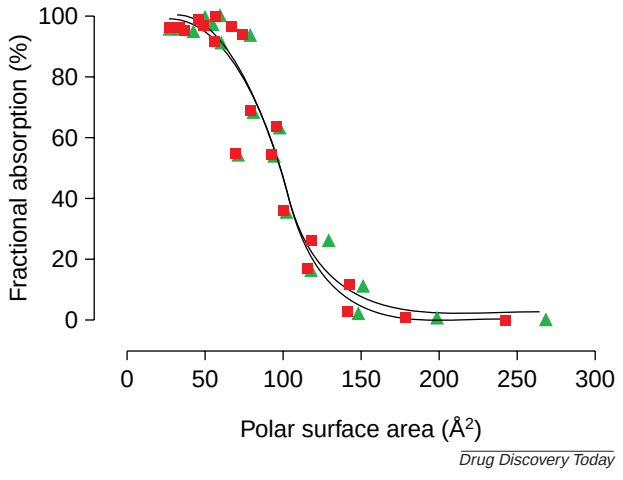

Figure 3. *Correlation between human fractional absorption and polar surface area calculated using the 'dynamic' (in red)[27] and single conformer methods (in green)[31].*

Prediction of BBB penetration

The BBB is a complex cellular system whose purpose is to maintain the homeostasis of the CNS by separating the brain from the systemic blood circulation[36]. For drugs targeted at the CNS, BBB penetration is a necessity (unless invasive or intranasal delivery routes are being considered). By contrast, for drugs aimed at other sites of action, passage through the BBB might produce unwanted side-effects. Hence, in drug discovery, the determination of BBB penetration is of great importance, and *in vitro*[37] and *in vivo*[38] techniques have been devised for this purpose. Computational approaches to the prediction of BBB penetration have been recently reviewed[39]. For the purposes of this review, they can be divided into two classes: those based on observations of CNS activity/inactivity and those seeking to model a specific experimental measure of BBB penetration.

Predictions based on CNS activity/inactivity

Initial work examining CNS activity[40] used a data set of 28 compounds, 16 of which showed high CNS activity (class designated CNS^+) with the remainder showing low or no CNS activity (CNS^-). Using discriminant analysis, three different models were derived, each using a different combination of topological descriptors together with a hydrogen-bonding parameter. Interestingly, although ClogP was available as a variable, it was not incorporated into any of the models during the stepwise selection process. Each of the models performed well, classifying all the CNS^+ compounds correctly, and only one of the CNS^- set incorrectly. Other workers[33] have examined a larger set of

Table 2. The 13 passively absorbed compounds and their $logP_{eff}$ values[32]

Compound	[a]$LogP_{eff}$	Compound	$LogP_{eff}$
Antipyrine	−3.35	Desipramine	−3.36
Hydrochlorothiazide	−5.40	Enalaprilat	−4.70
Verapamil	−3.17	Fluvastatin	−3.62
Ketoprofen	−3.08	Furosemide	−5.30
Metoprolol	−3.89	Naproxen	−3.08
Terbutaline	−4.52	Propanolol	−3.54
Carbamazepine	−3.37		

[a] Effective permeability.

CNS^+ and CNS^- compounds and concluded that, for brain penetration, MW should be <450 and PSA <90 Å^2.

More recently, the Vertex drug/non-drug discrimination approach has been applied to sets of CNS active/inactive compounds to bias the design of combinatorial libraries towards CNS targets [Ajay (1999) Biologically active molecules that cross the BBB: Designing combinatorial libraries with CNS activity. *Gordon Research Conference on Quantitative Structure–Activity Relationships*, 25–30 July, Tilton, NH, USA]. This kind of approach can take advantage of the large quantity of data available concerning CNS (in)activity. However, care must be taken to ensure that in the CNS^+ class, activity is caused by the parent compound and not by a metabolite. The CNS^- class is further complicated because CNS inactivity might not be solely caused by poor BBB penetration, as the compound might undergo extensive first-pass metabolism[41] or it might penetrate the brain but not exhibit any receptor binding[33]. In addition, as discussed in the section on absorption prediction, this method only supplies a binary answer, which is of limited use in lead optimization.

Predicting logBB

The second computational approach has been based mainly on the prediction of a brain penetration measure termed logBB, which is the ratio of the steady-state concentrations of the drug molecule between the brain and the blood [i.e. log (C_{brain}/C_{blood})]. Published values of logBB range from approximately −2.00 to +1.00. Within this range, compounds with logBB >0.3 cross the BBB readily, while those with logBB <−1.0 are only poorly distributed to the brain[42]. Several groups have employed QSPR approaches to develop models for the prediction of logBB.

An early attempt at predicting logBB used a small data set of 20 compounds to derive an equation relating logBB to PSA and molecular volume[43]. As with intestinal absorption, a reduced PSA favoured BBB penetration. Subsequently, however, it was demonstrated that this training set was too small to provide a robust predictive model for logBB (Ref. 44). Thus, Abraham and coworkers collated a larger set of approximately 60 compounds with logBB data that they successfully correlated with five solute descriptors[45]. Their results indicated that increasing molecular volume and excess molar refraction favours partitioning into the brain, while properties such as dipolarity/polarizability and hydrogen-bond acidity/basicity work against it. One disadvantage of the Abraham approach, as originally described, is that it requires manual dissection of the compound of interest and summation of descriptor values for the resulting fragments. This means it is unsuitable for use in computational screening of large compound collections (virtual HTS). Automation of the calculation of descriptor values has recently been reported[46], which should help to alleviate this problem.

The Abraham data set has formed the basis for several subsequent studies. Workers at Pfizer developed a QSPR correlating logBB with ΔG^{o}_{W} (the computed free energy of solvation of a compound in water)[47]. Norinder and coworkers have again used their MolSurf descriptors, in conjunction with PLS analysis, and concluded that partitioning into the brain is favoured by minimizing the number of groups capable of hydrogen bonding and by increasing lipophilicity[48]. However, both of these approaches require computationally expensive and non-automated calculation procedures. More rapid approaches have recently been reported[49,50], which offer the possibility of rapidly screening large (virtual) compound collections for their potential to cross the BBB. In the first of these more rapid approaches[49], many topological descriptors are analyzed by a PLS procedure, resulting in an 18-parameter equation that would be difficult to interpret for lead optimization. By contrast, a simple two-variable equation (r = 0.887) has been reported, derived from 55 compounds[50]:

$$\log BB = -0.0148 \times PSA + 0.152 \times ClogP + 0.139 \qquad (3)$$

Here, PSA is the single-conformer PSA (Ref. 31) and ClogP is the calculated octanol-water partition coefficient obtained from the Daylight ClogP software mentioned earlier. There are several advantages to this method for logBB prediction. Calculation is very rapid and is fully automated, requiring no human intervention. Furthermore, the quantities that comprise this equation are readily interpretable, making it simple for a medicinal chemist

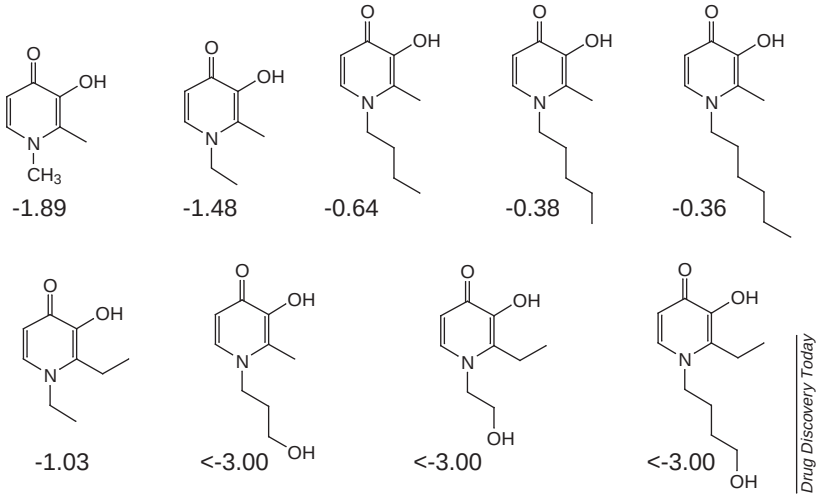

Figure 4. *Hydroxypyridinones with their logPS values.*

to alter the molecular structure to change the PSA and/or ClogP values, thereby changing the predicted logBB of a (hypothetical) compound. Equation (3) has demonstrated good predictive ability on several test sets[50,51].

Other measures of brain penetration

Whereas logBB is a measure of the steady-state distribution of a compound between the brain and the blood, logPS gives an indication of the permeability of the BBB to a compound[52]. LogPS is measured using a short-duration vascular perfusion method from which a permeability-surface area product is calculated (hence logPS). LogPS values for a set of 18 compounds have been correlated with Abraham's solute descriptors[52]. Further logPS data for nine hydroxypyridinones (Fig. 4) have recently been reported[53]. As more of these data emerge, together with data from *in vivo* techniques such as microdialysis, it will be interesting to see how they, like logBB, can be used to derive useful, predictive computational models.

Applications

It was recently predicted that, in less than five years, algorithms for the accurate prediction of passive permeation characteristics will be developed and available for routine use by medicinal chemists [Borchardt, R.T. (1999) Profiling compounds for biopharmaceutical properties: Overview. *Abstracts of the 217th American Chemical Society Meeting*, 21–25 March, Anaheim, CA, USA]. The previous sections have shown that good progress has been made towards this goal in recent years and it seems that our understanding of the molecular determinants of absorption is growing, leading to improved predictive models. Of course, such models are not simply of academic interest – they can play a vital role in focusing and accelerating drug discovery. However, it is fair to say that, at this time, most publications in this field have been methodological and there are few publications presenting successes (or failures) of such methods in an industrial setting. This section will discuss some brief indications of how these types of techniques are being applied in drug discovery.

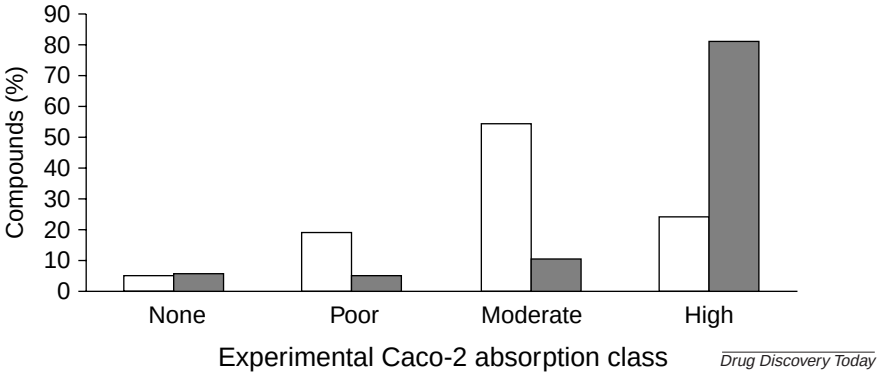

Figure 5. *Caco-2 absorption ratings for the two combinatorial libraries, LIB1 (open bars) and LIB2 (shaded bars), showing the improvement in absorption resulting from designing a library subject to polar surface area and rule-of-five constraints. The definitions of the classes are as follows: poor (<2% absorbed), moderate (2–20% absorbed) and high (>20% absorbed). The percentage absorbed figure is calculated by simply taking the peak area for the acceptor chamber and dividing it by the value for the donor chamber. The surface area of the inserts is 0.33 cm² (Costar HTS 24-well plate). The value shown is the mean taken from two experiments, each run over 3 h.*

Compound and screening set selection

One important application of these techniques is in the context of compound and screening set selection. For example, techniques such as those already described can be used

to filter a set of compounds from an external supplier prior to purchase. Workers at GlaxoWellcome[9] have applied the genetic algorithm-based method described earlier[10] to the selection of compounds from the corporate database for generation of a screening set. Profiling using the rule-of-five and PSA criteria can also provide a valuable indicator of the likely absorption characteristics of a combinatorial library or screening set.

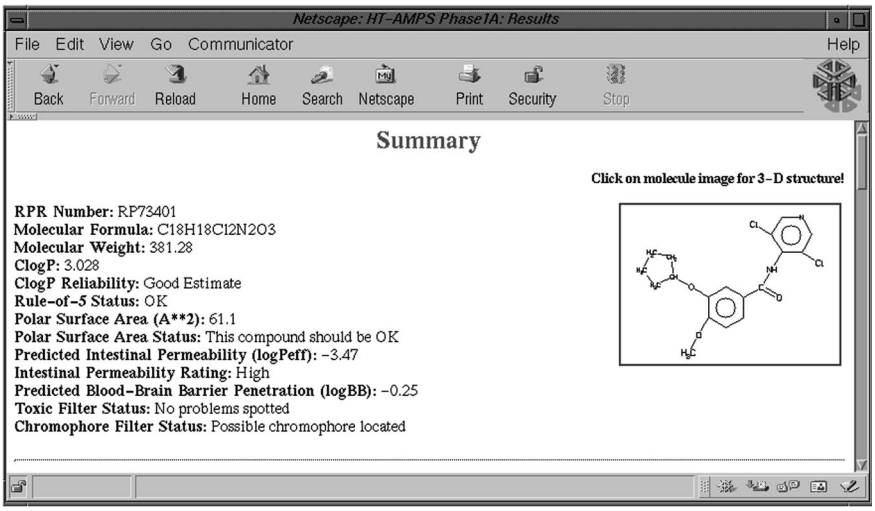

Figure 6. *Summary results from molecular property calculations submitted via a Web-based interface in use at RPR.*

Combinatorial library design

In addition to simply profiling libraries, this approach has been taken one stage further and has been applied to the design of combinatorial libraries[51]. In one example, a chemist had selected reagents for a combinatorial library (LIB1) in an oral drug discovery program to optimize parameters such as MW and ClogP in an approximate manner. A follow-up library (LIB2) was designed to optimize PSA and rule-of-five criteria much more rigorously with reagent selection being performed by a Monte Carlo search procedure[51]. Both libraries were subsequently tested in a Caco-2 monolayer absorption system and the designed library showed much improved absorption (Fig. 5). These results show the added value of considering quantities such as PSA in compound (library) design in addition to more traditional computed descriptors such as ClogP and MW.

Intranet-based systems

The speed and ease with which many molecular properties can be calculated means that such computations can be readily incorporated into an Intranet system for easy access by medicinal chemists and other researchers. In the system developed at Rhône-Poulenc Rorer (Dagenham, Essex, UK), the user can input a structure using a Java-based sketcher[54], a SMILES (simplified molecular input and line entry system) code[55] or, for known compounds, a corporate registry number. In a few seconds, the results are returned including values for simple molecular properties, predictions of absorption based upon the rule-of-five and PSA, predicted BBB penetration, as well as alerts for the presence of potentially toxic functionality (Fig. 6). Batch calculations using sets of SMILES codes are also possible via the Intranet. These computations are now used routinely by bench scientists. Other workers have reported similar approaches [O'Donnell, T.J. and Doman, T. (1999) ChemMart: One-stop Web shopping of structural databases. *Abstracts of the 217th American Chemical Society Meeting*, 21–25 March, Anaheim, CA, USA] and it is likely that this type of system will be commonly used in the industry in the future.

Conclusions

The rule-of-five was originally derived because of the realization that HTS was identifying large numbers of hit compounds, many of which did not possess 'drug-like' properties. It is clear from the growing number of publications in this area that this message has been taken on board, and methods for predicting 'drug-likeness' are already having a major impact on the design and selection of compounds in several pharmaceutical companies. The routine use of experimental absorption systems in the pre-screening of compounds is providing a wealth of data and should allow for the derivation of improved models for drug absorption. The true worth of such experimental and theoretical systems will become apparent when a reduction is seen in lead optimization times and attrition rates in preclinical and clinical phases of drug development.

Note added in proof

Readers are directed to the following references that are of relevance to the topic of predicting drug-likeness and that have appeared while this review was in press:

- Kelder, J. *et al.* (1999) Polar molecular surface area as a dominating determinant for oral absorption and brain penetration of drugs. *Pharm. Res.* 16, 1514–1519

- Stenberg, P. *et al.* (1999) Prediction of intestinal absorption of endothelin receptor antagonists using three theoretical methods of increasing complexity. *Pharm. Res.* 16, 1520–1526
- Wang, J. and Ramnarayan, K. (1999) Toward designing drug-like libraries: A novel computational approach for prediction of drug feasibility of compounds. *J. Comb. Chem.* 1, 524–533
- Ajay *et al.* (1999) Designing libraries with CNS activity. *J. Med. Chem.* 42, 4942–4951

REFERENCES

1 Smith, D.A. and van de Waterbeemd, H. (1999) Pharmacokinetics and metabolism in early drug discovery. *Curr. Opin. Chem. Biol.* 3, 373–378

2 Delie, F. and Rubas, W.A. (1997) A human colonic cell line sharing similarities with enterocytes as a model to examine oral absorption: Advantages and limitations of the Caco-2 model. *Crit. Rev. Ther. Drug Carrier Syst.* 14, 221–286

3 Irvine, J. *et al.* (1999) MDCK (Madin–Darby canine kidney) cells: A tool for membrane permeability screening. *J. Pharm. Sci.* 88, 28–33

4 Eddershaw, P.J. and Dickins, M. (1999) Advances in *in vitro* drug metabolism screening. *Pharm. Sci. Technol. Today* 2, 13–19

5 Walters, W.P. *et al.* (1999) Recognizing molecules with drug-like properties. *Curr. Opin. Chem. Biol.* 3, 384–387

6 Lewis, R.A. *et al.* (1997) Similarity measures for rational set selection and analysis of combinatorial libraries: The diverse property-derived (DPD) approach. *J. Chem. Inf. Comput. Sci.* 37, 599–614

7 Rishton, G.M. (1997) Reactive compounds and *in vitro* false positives in HTS. *Drug Discovery Today* 2, 382–384

8 Walters, W.P. *et al.* (1998) Virtual screening – An overview. *Drug Discovery Today* 3, 160–178

9 Hann, M. *et al.* (1999) Strategic pooling of compounds for high-throughput screening. *J. Chem. Inf. Comput. Sci.* 39, 897–902

10 Gillet, V.J. *et al.* (1998) Identification of biological activity profiles using substructural analysis and genetic algorithms. *J. Chem. Inf. Comput. Sci.* 38, 165–179

11 Ajay *et al.* (1998) Can we learn to distinguish between drug-like and non-drug-like molecules? *J. Med. Chem.* 41, 3314–3324

12 Sadowski, J. and Kubinyi, H. (1998) A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* 41, 3325–3329

13 Polinsky, A. (1999) Combichem and cheminformatics. *Curr. Opin. Drug Discovery Dev.* 2, 197–203

14 Navia, M.A. and Chaturvedi, P.R. (1996) Design principles for orally bioavailable drugs. *Drug Discovery Today* 1, 179–189

15 Chan, O.H. and Stewart, B.H. (1996) Physicochemical and drug-delivery considerations for oral drug bioavailability. *Drug Discovery Today* 1, 461–473

16 Krämer, S.D. (1999) Absorption prediction from physicochemical parameters. *Pharm. Sci. Technol. Today* 2, 373–380; 2, 426 (corrigendum)

17 Lipinski, C.A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23, 3–25

18 Moriguchi, I. *et al.* (1992) Simple method of calculating octanol/water partition coefficient. *Chem. Pharm. Bull.* 40, 127–130

19 Ghose, A.K. *et al.* (1999) A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* 1, 55–68

20 van de Waterbeemd, H. *et al.* (1996) Estimation of Caco-2 cell permeability using calculated molecular descriptors. *Quant. Struct.-Act. Relatsh.* 15, 480–490

21 Norinder, U. *et al.* (1997) Theoretical calculation and prediction of Caco-2 cell permeability using MolSurf parametrization and PLS statistics. *Pharm. Res.* 14, 1786–1791

22 Conradi, R.A. *et al.* (1992) The influence of peptide structure on transport across Caco-2 cells. II. Peptide bond modification which results in improved permeability. *Pharm. Res.* 9, 435–439

23 Palm, K. *et al.* (1996) Correlation of drug absorption with molecular surface properties. *J. Pharm. Sci.* 85, 32–39

24 Krarup, L.H. *et al.* (1998) Predicting drug absorption from molecular surface properties based on molecular dynamics simulations. *Pharm. Res.* 15, 972–978

25 Palm, K. *et al.* (1998) Evaluation of dynamic polar molecular surface area as predictor of drug absorption: Comparison with other computational and experimental predictors. *J. Med. Chem.* 41, 5382–5392

26 Stenberg, P. *et al.* (1999) Prediction of membrane permeability to peptides from calculated dynamic molecular surface properties. *Pharm. Res.* 16, 205–212

27 Palm, K. *et al.* (1997) Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm. Res.* 14, 568–571

28 Norinder, U. *et al.* (1999) Theoretical calculation and prediction of intestinal absorption of drugs in humans using MolSurf parametrization and PLS statistics. *Eur. J. Pharm. Sci.* 8, 49–56

29 Ghuloum, A.M. *et al.* (1999) Molecular hashkeys: A novel method for molecular characterization and its application for predicting important pharmaceutical properties of molecules. *J. Med. Chem.* 42, 1739–1748

30 Wessel, M.D. *et al.* (1998) Prediction of human intestinal absorption of drug compounds from molecular structure. *J. Chem. Inf. Comput. Sci.* 38, 726–735

31 Clark, D.E. (1999) Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. *J. Pharm. Sci.* 88, 807–814

32 Winiwarter, S. *et al.* (1998) Correlation of human jejunal permeability (*in vivo*) of drugs with experimentally and theoretically derived parameters. A multivariate data analysis approach. *J. Med. Chem.* 41, 4939–4949

33 van de Waterbeemd, H. *et al.* (1998) Estimation of blood–brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors. *J. Drug Targeting* 6, 151–165

34 Goodwin, J.T. *et al.* (1999) Strategies toward predicting peptide cellular permeability from computed molecular descriptors. *J. Pept. Res.* 53, 355–369

35 Raevsky, O.A. and Schaper, K-J. (1998) Quantitative estimation of hydrogen bond contribution to permeability and absorption processes of some chemicals and drugs. *Eur. J. Med. Chem.* 33, 799–807

36 Pardridge, W.M. (1998) CNS drug design based on principles of blood–brain barrier transport. *J. Neurochem.* 70, 1781–1792

37 Cecchelli, R. *et al.* (1999) *In vitro* model for evaluating drug transport across the blood–brain barrier. *Adv. Drug Deliv. Rev.* 36, 165–178

38 de Lange, E.C.M. *et al.* (1999) Microdialysis for pharmacokinetic analysis of drug transport to the brain. *Adv. Drug Deliv. Rev.* 36, 211–227

39 George, A. (1999) The design and molecular modeling of CNS drugs. *Curr. Opin. Drug Discovery Dev.* 2, 286–292

40 Basak, S.C. *et al.* (1996) Predicting blood–brain transport of drugs: A computational approach. *Pharm. Res.* 13, 775–778

41 Fischer, H. *et al.* (1998) Blood–brain barrier permeation: Molecular parameters governing passive diffusion. *J. Membr. Biol.* 165, 201–211

42 Abraham, M.H. *et al.* (1997) On the partition of ampholytes: Application to blood–brain distribution. *J. Pharm. Sci.* 86, 310–315

43 van de Waterbeemd, H. and Kansy, M. (1992) Hydrogen-bonding capacity and brain penetration. *Chimia* 46, 299–303

44 Calder, J.A. and Ganellin, C.R. (1994) Predicting the brain-penetrating capability of histaminergic compounds. *Drug Des. Discov.* 11, 259–268

45 Abraham, M.H. *et al.* (1994) Hydrogen bonding. 33. Factors that influence the distribution of solutes between blood and brain. *J. Pharm. Sci.* 83, 1257–1268

46 Platts, J.A. *et al.* (1999) Estimation of molecular free energy relation descriptors using a group contribution approach. *J. Chem. Inf. Comput. Sci.* 39, 835–845

47 Lombardo, F. *et al.* (1996) Computation of brain–blood partitioning of organic solutes via free energy calculations. *J. Med. Chem.* 39, 4750–4755

48 Norinder, U. *et al.* (1998) Theoretical calculation and prediction of brain–blood partitioning of organic solutes using MolSurf parametrization and PLS statistics. *J. Pharm. Sci.* 87, 952–959

49 Luco, J.M. (1999) Prediction of the brain–blood distribution of a large set of drugs from structurally derived descriptors using partial least-squares (PLS) modeling. *J. Chem. Inf. Comput. Sci.* 39, 396–404

50 Clark, D.E. (1999) 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.* 88, 815–821

51 Pickett, S.D. *et al.* Enhancing the hit-to-lead properties of lead optimization libraries. *J. Chem. Inf. Comput. Sci.* (in press)

52 Gratton, J.A. *et al.* (1997) Molecular factors influencing drug transfer across the blood–brain barrier. *J. Pharm. Pharmacol.* 49, 1211–1216

53 Habgood, M.D. *et al.* (1999) Investigation into the correlation between the structure of hydroxypyridinones and blood–brain barrier permeability. *Biochem. Pharmacol.* 57, 1305–1310

54 Ertl, P. (1998) World Wide Web-based system for the calculation of substituent parameters and substituent similarity searches. *J. Mol. Graph. Model.* 16, 11–13

55 Weininger, D. (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* 28, 31–36

Collaboration…

Cambridge Drug Discovery (Cambridge, UK) has engaged in a collaborative agreement with **KuDOS Pharmaceuticals** (Cambridge, UK) for high-throughput screening (HTS) to identify lead compounds as cancer therapies. As part of the agreement, both companies will work together to identify novel inhibitors of enzymes involved in DNA repair, while KuDOS will gain access to Cambridge Drug Discovery's fully automated screening facilities. Barry Kenny, the Research Director of Cambridge Drug Discovery, said, 'Integration of assay development into a quality HTS environment is a key activity for us and we are confident of providing exciting leads on this programme.'

Contributions to *Drug Discovery Today*

Drug Discovery Today publishes topical information on all aspects of drug discovery – molecular targets, lead identification, lead optimization and associated technologies – together with overviews of the current status of compound classes, approaches in specific therapeutic areas or disease states and novel strategies, such as gene therapy. Areas of pharmaceutical development that relate to the potential and viability of drug candidates are also included, as are those relating to the strategic, organizational and logistic issues underlying pharmaceutical R&D.

Authors should aim for topicality rather than comprehensive coverage. Ultimately, articles should improve the reader's understanding of the field addressed and should therefore assist in the increasingly important decision-making processes for which drug discovery scientists are responsible.

Most articles appearing in *Drug Discovery Today* are commissioned. However, suggestions and proposals for full reviews or shorter items for the *Editorial*, *Monitor* or *Update* sections are welcomed; in the first instance, a tentative title and brief outline of the proposed article should be supplied. Typically, full reviews will extend to 4000 words with up to 60 references. *Update* and *Monitor* items (news and views, reports of new technological advances, conferences, experimental methods, and critical assessment of important new literature and other media) do not usually exceed 1000 words, and one or two figures plus up to ten references may be included. The *Editorial* represents a personal perspective on contemporary issues and controversies affecting R&D and the pharmaceutical industry.

If you would like to contribute to *Drug Discovery Today* in future, please submit your proposal to: Dr Debbie Tranter, Editor, *Drug Discovery Today*, Elsevier Science London, 84 Theobald's Road, London, UK WC1X 8RR.