DDT Vol. 7, No. 20 October 2002

Application of hydrogen bonding calculations in property based drug design

Michael H. Abraham, Adam Ibrahim, Andreas M. Zissimos, Yuan H. Zhao, John Comer and Derek P. Reynolds

A drug can be characterized by 'descriptors' that include size (volume) and H-bond acidity and H-bond basicity. These descriptors can be rapidly estimated from structure by a fragment scheme and used to predict physicochemical and transport properties of drug candidates (e.g. log*P*, solubility, gastrointestinal absorption, permeability and blood–brain distribution). The solvation equations can be interpreted to provide a qualitative chemical insight into biological partition and transport mechanisms. Applications to blood–brain partition and human intestinal absorption (HIA) are discussed.

Michael H. Abraham
Adam Ibrahim
Andreas M. Zissimos
Yuan H. Zhao
Dept of Chemistry
University College London
20 Gordon Street
London, UK WC1H OAJ
John Comer
Sirius Analytical Instruments
Riverside
Forest Row Business Park
Forest Row
UK RH18 5DW
***Derek P. Reynolds**
REYTEK
11C Rothsay Road
Bedford, UK MK40 3PP
*e-mail: reytek@ntlworld.com

▼ It is a major challenge to convert a compound binding with high affinity to a biological target (i.e. a hit, lead or candidate molecule) into a successful drug on the market. Attrition during the drug development process is a serious economic problem for the pharmaceutical industry and it is often a result of inappropriate physicochemical characteristics and related poor absorption and poor pharmacokinetics [1]. It is now widely recognized that the early drug discovery process needs to address in parallel not only potency (in relation to the primary target), but also other biological (selectivity), toxicological and physicochemical (pharmacokinetics related) properties [2]. Van de Waterbeemd and his colleagues at Pfizer have called this approach 'property-based design' [3] and it emphasizes the importance of understanding how medicinal chemists can manipulate the crucial combinations of physical and structural properties which contribute to 'drug-likeness'.

Drug solubility and the ability of molecules to permeate barriers are two key properties that are often considered to be linked to lipophilicity. Excessive lipophilicity is a common cause of poor solubility and can lead to erratic and incomplete absorption following oral administration. A biopharmaceutics classification scheme has been proposed under which drugs can be categorized into four groups according to their solubility and permeability properties [4]. However it has been argued that the principal properties are not solubility and permeability, but rather the basic components of both, namely molecular size and hydrogen (H)-bonding [5]. Octanol–water partition (log*P*) and distribution (log*D*) coefficients are the most widely used measures of lipophilicity and several different schemes have been devised to provide a quantitative explanation of log*P* in terms of molecular size and polar parameters (e.g. H-bonding terms) [6–9].

In more recent years it has been found that processes involving passive diffusion (e.g. membrane permeation, intestinal absorption) depend primarily on the H-bonding capacity or polar surface area of a drug solute [10–14]. Quantitative experimental measures of H-bond strengths can be derived from partition measurements in different organic solvents, but recent drug discovery applications require high-throughput methods and computational approaches to H-bonding are generally used. Some calculations are simple and H-bonding contributions are crudely estimated from simple parameters such as the number of H-bond donors and acceptors. For example, the well-known Lipinski 'rule-of-five' generates an alert (indicating possible absorption problems) for compounds where any two of the following conditions is satisfied [15]:

1359-6446/02/$ – see front matter PII: S1359-6446(02)02478-9

- Molecular weight >500
- Number of H-bond donors >5
- Number of H-bond acceptors >10
- Octanol–water clog*P* >5

Clearly such a rule is only qualitative and for more quantitative predictions a better estimation of H-bond strength is required.

Choice of parameters to describe ‘drug-likeness’

Property distributions in drug databases have been studied as an approach to understanding ‘drug-likeness’ [16,17] and recently the difference between ‘drugs’ and ‘leads’ has also been examined [18,19]. These all show the importance of molecular size and H-bonding capacity. Computational methods have been used to estimate absorption through various barriers [20], for example, intestinal absorption [21–27], blood–brain transport [28–33], percutaneous absorption [34–37] and corneal permeability [38]. Both correlations with single properties (e.g. polar surface area [21,22,29,31]), as well as multiple variable approaches have been used. A problem for the medicinal chemist who wishes to investigate and understand the behaviour of a new lead series is that many different descriptors can be calculated using a range of commercially available software packages, such as Cerius2, Tsar, Hybot, QsarIS, and so on. Often when multivariate statistical approaches are applied it is some rather obscure or ill-defined descriptors that best describe a specific dataset. For example, Wessel *et al.* developed a six-descriptor non-linear computational neural network model for the estimation of %HIA (human intestinal absorption) for a dataset of 86 drugs [26]. The parameters used were clearly related to molecular size and H-bonding, but some of them, for example, SHDW-6 (normalized two-dimensional projection of molecule on XY plane), GRAV-3 (cube root of gravitational index) and SCAA-2 (surface area × charge of H-bond acceptor atoms/number of H-bond acceptor atoms) are for many chemists difficult to relate to specific structural elements or the properties of functional groups.

Approaches for estimating the strengths of H-bonds can be divided into two groups, quantum chemical calculations and empirical methods. The use of *ab initio* calculations allows the energy and geometric structures of small molecules to be described with an accuracy approaching those of experimental results. However, a similar treatment of large molecules is not yet practical. Reliable empirical H-bond parameters for drug-like molecules need to be derived from large databases of experimental measurements. Raevsky’s program HYBOT is linked to a database of several thousand thermodynamic measurements of 1:1 H-bonding systems and can estimate H-bond donor and acceptor factors for the relevant sites in any compound based on its chemical structure [39]. There is however a difference between H-bonding scales that refer to 1:1 complexation and scales that refer to the overall or summation H-bond acidity and basicity. The second scales deal with the situation in which a solute is surrounded by solvent molecules so that all acid (i.e. H-bond donor) groups and all basic (i.e. H-bond acceptor) groups in the solute are potentially involved in acid–base interactions at the same time. In the approach by Abraham [8] these differences are explicitly addressed and it has been concluded [40] that functional group constants derived from overall H-bond parameters are the more useful in analyses of physicochemical and biochemical properties.

Property-based design [3] is a difficult problem that requires, on the one hand, an understanding of the properties of biological barriers and, on the other hand, an adequate and consistent description of the properties of drug molecules. Simple parameters (e.g. number of H-bond donor and acceptor sites) are highly useful but we believe that a more rigorous approach is desirable for the description of molecules and advocate the use of Abraham descriptors and the general solvation equation. In the following sections this approach is explained in more detail and some applications are examined.

Abraham descriptors and the solvation equation

The Abraham approach based on the solvation equation correlates solute properties such as solubility, partitioning, blood–brain distribution, cell permeability, and human intestinal absorption, with a set of five molecular descriptors. These solute descriptors are based on the physically meaningful theoretical cavity model of solute–solvent interactions (Fig. 1). They are:

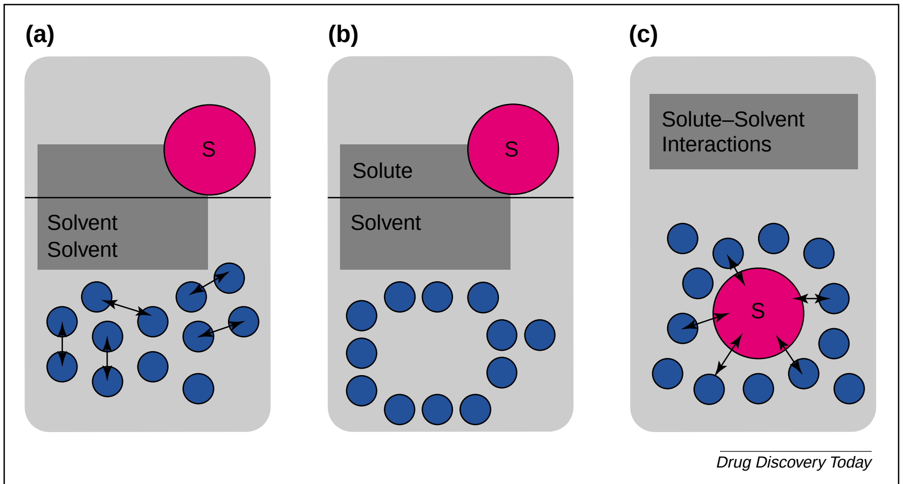
- *E* – excess molar refraction which models dispersion force interactions arising from the greater polarisability of pi- and n- electrons (E = 0 for saturated alkanes)
- *S* – solute polarity/polarisability (due to solute-solvent interactions between bond dipoles and induced dipoles)
- *A* – solute H-bond acidity
- *B* – solute H-bond basicity
- *V* – McGowan characteristic molar volume

The H-bonding parameters are summation terms relevant to the behaviour of solutes in solvents. The acidity ‘*A*’ relates to the strength and number of H-bonds formed by donor groups in the solute when they interact with lone pairs of acceptor groups in solvent molecules. The basicity ‘*B*’ relates to the strength and number of H-bonds formed by the lone pairs of acceptor groups in the solute when they interact with donor solvents. Although perhaps not familiar to all medicinal chemists, once understood, the Abraham descriptors are chemically intuitive.

The range of solutes for which descriptors are currently available is now quite large and encompasses compounds that range from simple gases (e.g. hydrogen, nitrogen) to complex drugs and pesticides (Table 1). For saturated hydrocarbons, *A*, *B*, *S* and *E* are all zero. The properties of compounds are therefore effectively referenced to the properties of saturated hydrocarbons. For most other compound classes the descriptors are positive, however, a few compound types, for example, perfluorinated alkanes have negative values for *S* and *E* and behave as less polar and more hydrophobic solutes than the corresponding hydrocarbons.

The solute descriptors can be combined in a linear free energy relationship (LFER; Eqn 1):

$$SP = c + eE + sS + aA + bB + vV \quad \text{[Eqn 1]}$$



Drug Discovery Today

Fig. 1. The transfer of a drug from one phase to another can be considered in terms of the difference in solvation of the drug in the two phases. The cavity model that we use breaks down solvation into three steps. **(a)** The intermolecular forces holding the phase (or solvent) together have to be broken to form a cavity of the same size as the drug (or solute) and this is energetically unfavorable. **(b)** The phase (or solvent) molecules around the cavity have to be reorganized. **(c)** Solute–solvent interactions are set up when the solute is inserted into the cavity and these are all energetically favorable. The work of forming the cavity in step (a) will be proportional to the drug volume, *V* and the general dispersion interaction in (c) is also proportional to *V* so that the *vV* term in the solvation equation includes a cavity effect, and an opposite general dispersion interaction effect. The various other solute–solvent interactions in step (c) are described by the terms *E*, *S*, *A* and *B*.

Where the dependent variable, SP, is a solute property in a given system. For example, it might be log*P*, for a set of solutes in a given water-solvent partition system. The coefficients in the equations are found by the method of multiple regression. Because the solute descriptors (*E*, *S*, *A*, *B* and *V*) represent the solute influence on various solute-solvent phase interactions, the regression coefficients *c*, *e*, *s*, *a*, *b* and *v* correspond to the complementary effect of the solvent phases on these interactions. The coefficients can then be regarded as system constants which characterize the phase and contain chemical information about the phase in question.

As an example, the important octanol–water system is characterised by Eqn 2 [9]:

$$\text{Log}P_{\text{oct}} = 0.088 + 0.562E - 1.054S - 0.032A - 3.460B + 3.814V$$
$$N = 613,\ R = 0.9974,\ \text{SD} = 0.116,\ F = 23161.6 \quad \text{[Eqn 2]}$$

The *a*-coefficient is almost zero, which shows that octanol is as strong a H-bond base as is water. However, water is a very strong H-bond acid and H-bond basic solutes (i.e. large *B*) interact less strongly with octanol resulting in a negative *b*-coefficient. Octanol is also less polar (i.e. dipolar/polarizable) than water hence the negative *s*-coefficient. Dispersion force interactions are taken into account by the terms involving *E* and *V*. The large positive *v*-coefficient is in part a result of the higher energy required to create a given-sized cavity in water, and in part a result of the greater dispersion force interactions with solutes in octanol. Solutes with π- and n-electron pairs are more polarizable (*E*>0) than the corresponding saturated hydrocarbons (*E* = 0) and this results in stronger dispersion interactions with octanol as shown by the positive *e*-coefficient.

In the Abraham QSAR approach the same five descriptors are always forced into the regression equation. Although in statistical terms there is sometimes little benefit in using more than four of the parameters, there is a big practical advantage because the equations for different properties can be directly compared. For example, a group from Pfizer

Table 1. Validated solute descriptors available in the UCL database

Descriptor	Maximum value	Minimum value	Total number available
A (H-bond acidity)	4.33	0.00	4490
B (H-bond basicity)	4.52	0.00	3440
S (Dipolarity/polarisability)	5.60	–0.54	3760
E (Excess molar refraction)	4.62	–1.37	4290
V (Molar volume)	8.56	0.07	4380

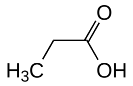
developed a high-throughput reverse-phase-HPLC method for determination of logP_{oct} values. A comparison of the derived solvation equations demonstrated that the HPLC method encoded the same information as obtained from a shake flask logP_{oct} determination [41]. In a similar but more general approach to lipophilicity determination the coefficients of solvation equations have been used to characterize the different selectivity characteristics of various HPLC stationary phases and organic modifiers [42].

Because they are derived from experimental data using a solvation model that accounts for all the major intermolecular interactions (i.e. dispersion forces, dipole–dipole, dipole-induced dipole and H-bonding), Abraham descriptors should provide a consistent and fairly complete set for the derivation of QSAR relationships that explain properties that are not sensitive to molecular shape or conformation (i.e. usually solution-phase properties that involve rapid molecular motions). It is desirable to be able to relate these empirical parameters to other purely theoretical descriptions of molecules and to be able to calculate Abraham parameters directly using various theoretical approaches. The various ways of doing this have recently been compared [43].

Methods for obtaining solute parameters

Any application of the general solvation equation (Eqn 1) depends on the availability of the solute descriptors, and the need to calculate descriptors for new compounds is of primary importance. The descriptor *V* can be calculated quite simply from the molecular formula and the number of rings in the molecule. The *E* descriptor can be calculated either from the observed or calculated liquid refractive index. The remaining three descriptors *S*, *A*, and *B* have to be obtained by analogy to other closely related compounds, by fragment addition, or by fitting to a set of experimental measurements such as log*P* values for a given solute in several different water–solvent systems [44]. Many alternative computed QSAR parameters are difficult to validate and the ability to derive Abraham descriptors from experimental data clearly differentiates them from such purely theoretical numbers. However, it is clear that any predictive application of the solvation equation (e.g. to guide library design before synthesis) requires rapid calculation methods. Platts and co-workers used the database of known values (see Table 1) to identify common substructures and relevant intramolecular interactions and evaluate their contributions to each descriptor. The final model used 81 fragment values for *E*, *S*, and *B* with a separate set of 51 fragments for calculation of *A*. Typically, errors of ~0.05–0.15 log unit (for values covering a range of 2–6 log units) were found [45].

Applications such as the high-throughput prediction of physicochemical and biochemical properties for compound libraries require access to suitable software. The published calculation scheme [45] for Abraham descriptors can be implemented using any suitable chemical software tools. For scientists that want a ready-made solution, the program *Absolv* is available for a Windows PC (Sirius Analytical Instruments; http://www.sirius-analytical.com). It has a straightforward graphical interface, can predict values for over 120 solute properties based on published LFER equations, and provides references to the original source literature. An example of the calculation of descriptors by *Absolv* is shown in Fig. 2, and a screenshot of the program is in Fig. 3.



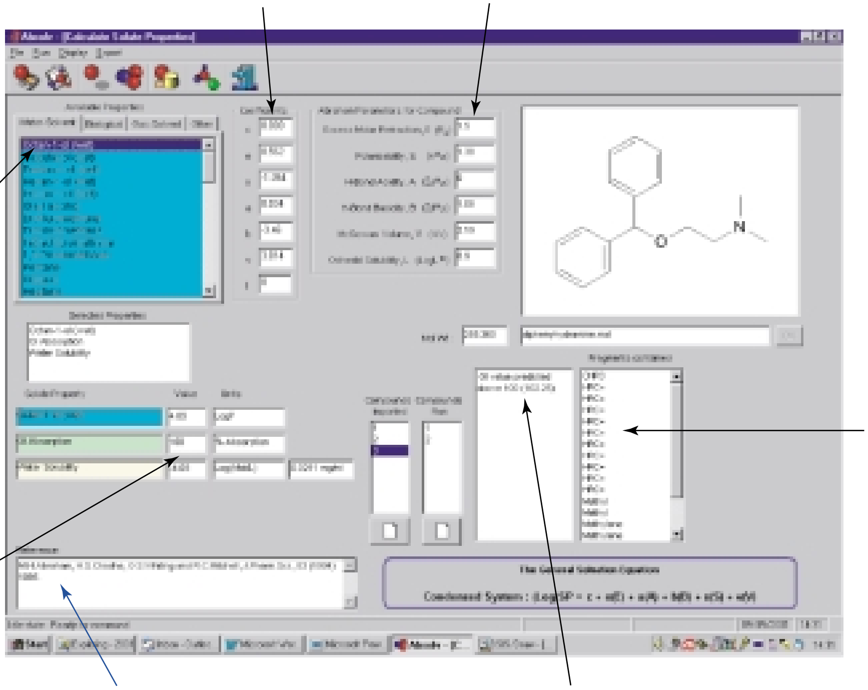
Fragment	A	B	S	E
Acid (-COOH), correction	0.26	-0.44	-0.40	-0.22
(-OH)	0.33	0.43	0.24	0.08
Methyl (-CH$_3$)	0.00	0.00	-0.08	-0.11
Methylene (-CH$_2$-)	0.00	0.00	0.00	0.00
Carbonyl (=O)	0.00	0.34	0.40	-0.06
R$_2$C= SP$_2$ Carbon	0.00	0.02	0.10	0.20
Intercept	0.00	0.07	0.27	0.25
Sum of group descriptors	0.59	0.41	0.53	0.14
Observed value	0.60	0.45	0.65	0.23

Drug Discovery Today

Fig. 2. The descriptors required for the application of the general solvation equation have been obtained from experimental data for some 3,500 drug and non-drug molecules. They can be calculated from structure by a fragment contribution method as included in the *Absolv* program (Sirius Analytical). Details for propanoic acid are shown above. *Absolv* is PC-based software that enables descriptors to be calculated for ~80–200 molecules per minute, starting from chemical structure. The calculation of numerous properties, such as % intestinal absorption, blood–brain distribution, skin permeation, and so on, can be carried out simultaneously at the same rate.

Applications of the solvation equation in property-based design

The solvation equation (Eqn 1) provides an investigational framework that allows relatively small sets (>30) of molecules to be used to probe the characteristics of biological transport systems. The resulting regression coefficients (*c*, *e*, *s*, *a*, *b* and *v*) encode chemical information about the barrier and the contributing equilibria and/or kinetic processes. Furthermore, the coefficients for different putative biological mimetic systems can be compared, to uncover significant mechanistic differences. The solute descriptors (*E*, *S*, *A*, *B*, and *V*) provide a consistent description of molecules and they comprise a generally useful set of QSAR parameters. Until recently, the lack of rapid approaches



Coefficients for the selected property

Abraham parameters calculated for the displayed molecule

Lists of 125 available properties.

Click here to highlight and select property

Up to five properties can be selected and appear here for calculation

Calculated results

Fragments of structure appear here. Individual fragments can be highlighted to display the Abraham Parameter contributions for that fragment

Literature reference shows where the selected equation was first published

Comments

Drug Discovery Today

Fig. 3. Screenshot from the *Absolv* program. It is an easy to use and commercially available Windows-based PC tool for calculating Abraham molecular descriptors from molecular structure and predicting the properties of molecules including drugs, agrochemicals and pollutants. It can predict values for 125 solute properties (e.g. log*P*, solubility in water, gastrointestinal absorption, permeability through skin, and blood–brain distribution) based on published linear free energy relationship (LFER) equations, and provides references to the original source literature.

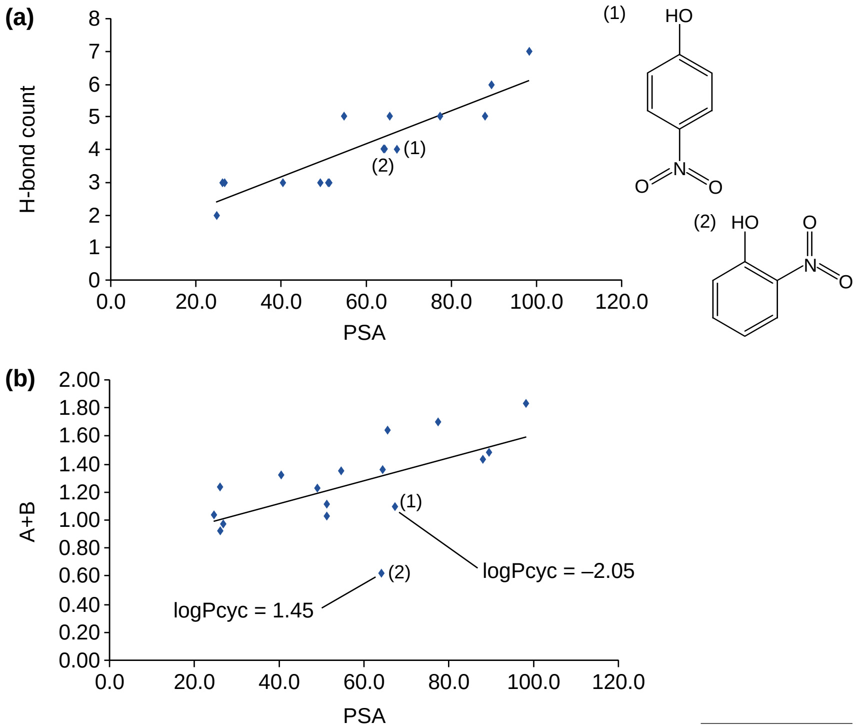
to parameter estimation from chemical structure has limited their application, but the availability of commercial software removes this restriction. However, it is important to appreciate that the paucity of reliable *in vivo* and *ex vivo* biological data is a primary restriction that limits our ability to properly understand the factors controlling absorption and distribution.

Brain penetration

Brain uptake of drugs is a subject of great importance to the pharmaceutical industry and the two measures that are most amenable to physicochemical analysis are: (a) blood–brain distribution and (b) rate of uptake during brain perfusion. Most work has been done on blood–brain distribution (BB), and several workers [29,30,31] have analyzed a dataset similar to that of 57 compounds first assembled by Abraham and co-workers [46]. Recently the set has been extended to 148 compounds [33] and the solvation equation determined (Eqn 3):

$$\text{LogBB} = 0.044 + 0.511E - 0.886S - 0.724A - 0.666B + 0.861V$$
$$N = 148,\ R = 0.843,\ \text{SD} = 0.367,\ F = 71.0 \qquad \text{[Eqn 3]}$$

Reasonable statistics were obtained for Eqn 3, however several large discrepancies were observed for carboxylic acid containing molecules such as salicylic acid and indomethacin, and an improvement over Eqn 3 was obtained by inclusion of an indicator variable. The equations are significantly different from Eqn 2, which describes octanol–water log*P*, and show that size strongly enhances brain uptake, whereas polarity/polarizability, H-bond acidity, basicity, and the presence of carboxylic acid groups strongly reduce brain penetration. This model is as good as any of the other reported QSAR models that have used only subsets of the data. It is worth noting that several alternative models include polar surface area (PSA) [29,31] or topological polar surface area (TPSA) [47] as a descriptor of logBB with encouraging results. This is not in conflict with Eqn 3 because PSA correlates roughly with the sum $A + B$ but PSA and TPSA are more simplistic parameters that are not sensitive to H-bond strength and are not influenced by interactions such as intramolecular H-bonding as shown in Fig. 4. TPSA is calculated simply by adding contributions from 43 polar atom types and there are no interaction corrections, whereas the calculation routines for *A* and *B* will



Drug Discovery Today

Fig. 4. Polar surface area (PSA) has been used for the prediction of various drug transport properties [47]. Simple H-bond counts have also been used for fast computational screening and shown to correlate closely with PSA [14]. PSA and H-bond counts, are not sensitive to H-bond strengths and cannot reveal important details such as internal H-bonding. In the example shown, two compounds, *ortho*- and *para*-nitrophenol are indistinguishable in PSA terms despite the fact that the *ortho*- substituent exhibits an intra-molecular H-bonding interaction. This interaction is clearly shown to be important when the Abraham descriptors of the two molecules are considered. As a direct consequence of intra-molecular H-bonding the H-bond acidity of the phenolic hydrogen in the *ortho*- substituted phenol is very low ($A = 0.20$) compared with that of the para-substituted phenol ($A = 0.87$). The basicity of the *ortho*-substituted phenol ($B = 0.42$) is higher than the *para*-substituted phenol ($B = 0.22$). This difference between the two molecules is reflected in their properties. An example of this is the log$P$ value in the water–cyclohexane system. LogP_{cyc} for the *ortho*-phenol (1.45) is much larger than for the *para*-phenol (–2.05), even though they have the same H-bond count, and similar PSAs.

only work correctly for new compounds if the relevant interactions have been observed in the training set and properly parameterized. When compounds containing new fragments are synthesized it is always possible to determine the correct overall solute descriptors by making appropriate partition measurements.

Brain perfusion experiments measure a rate constant from experiments over a short timescale and because of the variation in experimental procedure the various sets of available data are not compatible. Analysis of rate constants (k) from one set of rat brain perfusion experiments [48] gave Eqn 4:

$$\text{Log k} = -1.213 + 0.711E - 1.868S - 0.0A - 2.795B + 3.301V$$
$$\text{N} = 18,\ \text{R} = 0.976,\ \text{SD} = 0.48,\ \text{F} = 65 \qquad \text{[Eqn 4]}$$

Eqn 4 shows clear similarities to octanol–water partition (Eqn 2) but is very different from Eqn 3 and demonstrates that the rate of transport through the blood–brain barrier should not be equated or confused with the equilibrium partition of drugs between blood and brain.

Oral absorption

Predictive methods for oral absorption are a first step towards the even more difficult problem of predicting human oral bioavailability. Absorption prediction is not simple because each cellular membrane can be considered as a both a physicochemical and a biological barrier to drug transport. Active transport mechanisms may overcome poor physicochemical properties and before any absorption can take place at all the first important process is often dissolution. Many cases of dissolution-rate limited absorption have been reported and so solubility is a property that needs to be considered in the early stages of drug discovery. After dissolution moderately lipophilic compounds will generally diffuse through the lipid membranes of the gastrointestinal tract (transcellular absorption) whereas small hydrophilic compounds diffuse through the aqueous pores at the tight junctions between cells (paracellular absorption). Because animals have evolved mechanisms to protect them against harmful xenobiotics, the membranes of the gut wall contain metabolizing enzymes (e.g. CYP3A4) and transporter proteins (e.g. P-glycoprotein) that can act together to reduce the absorption of any compounds that are substrates for their rather indiscriminate binding sites. To have good oral absorption a drug must have a high transmembrane flux as well as avoid efflux by P-glycoprotein and metabolism by gut wall CYP3A4. As a pragmatic approach to designing orally absorbed drugs, it has been suggested that focussing on the factors that give a drug a high permeation rate will give the best chance of saturating the efflux component and the metabolizing enzymes of the gut wall [3].

As noted previously, the paucity of suitable data often limits our ability to construct reliable QSAR. For human intestinal absorption a large dataset of compounds was rigorously analyzed by Zhao *et al.* [27]. Data were collected from the literature on 241 drugs and critically appraised and classified so that they could be analyzed with some confidence. Data from 169 drugs were suitable for inclusion in the analysis and were used to derive a solvation equation that directly related the percentage of drug absorbed to the Abraham descriptors E, S, A, B and V. Abraham and Le [49] have obtained a modified general solvation equation for aqueous solubility that includes a cross-term $A*B$, to take account of intermolecular interactions in liquids or solids, so that no melting point is needed. It is of interest that the H-bonding terms (A and B) that act to increase aqueous solubility act in the reverse way to decrease intestinal absorption. For comparison with other solvation equations an effective rate of absorption (GI k_{eff}) was calculated (for those drugs where absorption was not either 0 or 100%) to obtain Eqn 5 [50]:

$$\text{Log (GI } k_{eff}) = 0.544 - 0.025E + 0.141S - 0.409A - 0.514B + 0.204V$$

$$N = 127,\ R = 0.89,\ \text{SD} = 0.29,\ F = 94 \qquad \text{[Eqn 5]}$$

Reasonable statistics were obtained for Eqn 5; however, the regression coefficients bear no relation at all to those in the numerous equations that have been constructed for water–solvent partition equilibria (e.g. Eqn 2). By contrast, Eqn 5 is similar to equations that describe diffusion processes, and this strongly suggests that diffusion through an unstirred water or mucus layer together with interfacial (mucus–membrane) transfer is the rate-determining step in human intestinal absorption.

In summary, the Abraham descriptors are able to successfully predict absorption for a diverse set of drugs. The significant descriptors are the H-bonding terms A and B although the volume term can also contribute. Other authors had previously analyzed subsets of this data and correlated intestinal absorption with polar surface area (PSA) [21,22,31]. It has been suggested that PSA in some way describes the desolvation of a compound as it moves from an aqueous to a lipid environment [28]. The rule PSA >140 Å^2 has been proposed as a crude filter to identify poorly absorbed compounds before synthesis [22] and it has also been demonstrated that PSA can be modelled using H-bonding descriptors [14]. Similar criteria are also relevant to the more difficult problem of oral bioavailability. Based on data for over 1100 drug candidates measured at GlaxoSmithKline it has been suggested [51] that compounds which meet only two criteria of (1) 10 or fewer rotatable bonds and (2) PSA <140 Å^2 (or 12 or fewer H-bond donors and acceptors) will have a high probability of good oral bioavailability in the rat.

Decision trees

Any models that attempt to account for all the factors involved for the prediction of *in vivo* transport and distribution of drugs will be complex. However, if methods can be developed to identify the critical or rate-limiting steps within any overall process, then more precise and more predictive QSAR relationships will be possible. Recently, the software program, Algorithm Builder (from Pharma Algorithms; http://www.ap-algorithms.com), has become available and it allows computational filters incorporating complex decision trees to be programmed through a simple graphical interface [52]. For example, a human intestinal absorption (HIA) filter has been devised that generates an alert for compounds predicted to have low absorption (HIA<10%). The model was derived from an analysis of data from over 1000 compounds with experimental HIA values classified into a binary format (i.e. high, low). The decision-tree criteria are based on rules that include descriptors based on charge (e.g. compounds containing quaternary N^+ are not absorbed), size, log$P$, PSA, and H-bond acidity (A). The HIA filter can itself be incorporated into other decision trees with additional rules; for example, to classify compounds according to their oral bioavailablility. These decision trees are easily updated by independent experts with new (custom) rules and have been termed ‘open-concept’ filters. Hopefully pharmaceutical companies will continue to publish new rules based on their in-house data and reveal the principles on which they are based, even when, as a result of the proprietary nature of their compounds, it becomes impossible to reveal the actual chemical structures from which the criteria were derived.

References

1 Hodgson, J. (2001) ADMET- turning chemicals into drugs. *Nat. Biotechnol.* 19, 722–726
2 Baxter, A.D. and Lockey, P.M. (2001) ‘Hit’ to ‘lead’ and ‘lead’ to ‘candidate’ optimisation using multi-parametric principles. *Drug Discov. World* 3, 9–15
3 Van de Waterbeemd, H. *et al.* (2001) Property-based design: optimisation of drug absorption and pharmacokinetics. *J. Med. Chem.* 44, 1313–1333
4 Amidon, G.L. *et al.* (1995) Theoretical basis for a biopharmaceutic drug classification: the correlation of *in vitro* drug product dissolution and *in vivo* bioavailability. *Pharm. Res.* 12, 413–420
5 Van de Waterbeemd, H. (1998) The fundamental variables of the biopharmaceutics classification system (BCS): a commentary. *Eur. J. Pharm. Sci.* 7, 1–3
6 Van de Waterbeemd, H. and Testa, B. (1987) The parametrization of lipophilicity and other properties in drug design. *Adv. Drug Res.* 16, 85–225
7 El-Tayar, A. *et al.* (1992) Polar intermolecular interactions encoded in partition coefficients: an indirect estimation of hydrogen-bond parameters of polyfunctional solutes. *J. Phys. Chem.* 96, 1455–1459
8 Abraham, M.H. (1993) Scales of hydrogen bonding – their construction and application to physicochemical and biochemical processes. *Chem. Soc. Rev.* 22, 73–83

9 Platts, J.A. *et al.* (2000) Estimation of molecular linear free energy relationship descriptors by a group contribution approach. 2. Prediction of partition coefficients. *J. Chem. Inf. Comput. Sci.* 40, 71–80
10 Clark, D.E. and Pickett, S.D. (2000) Computational methods for the prediction of 'drug-likeness'. *Drug Discov. Today* 5, 49–58
11 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
12 Zhao, Y.H. *et al.* (2001) Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure activity relationship (QSAR) with the Abraham descriptors. *J. Pharm. Sci.* 90, 749–784
13 Platts, J.A. *et al.* (2000) Estimation of molecular linear free energy relationship descriptors. 4. Correlation and prediction of cell permeation. *Pharm. Res.* 17, 1013–1018
14 Osterberg, T. and Norinder, U. (2000) Prediction of polar surface area and drug transport processes using simple parameters and PLS statistics. *J. Chem. Inf. Comput. Sci.* 40, 1408–1411
15 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
16 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
17 Oprea, T.L. (2000) Property distribution of drug-related chemical databases. *J. Comput. Aided Mol. Des.* 14, 251–264
18 Hann, M.M. *et al.* (2001) Molecular complexity and its impact on the probability of finding leads for drugdiscovery. *J. Chem. Inf. Comput. Sci.* 41, 856–864
19 Oprea, T.I. *et al.* (2001) Is There a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comput. Sci.* 41, 1308–1315
20 Norinder, U. and Osterberg, T. (2000) The applicability of computational chemistry in the evaluation and prediction of drug transport properties. *Perspect. Drug Discov. Des.* 19, 1–18
21 Palm, K. *et al.* (1998) Evaluation of dynamic polar molecular surface area as a predictor of drug absorption: comparison with other computational and experimental properties. *J. Med. Chem.* 41, 5382–5392
22 Clark, D.E. (1999) Rapid evaluation of polar surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. *J. Pharm. Sci.* 88, 807–814
23 Winiwater, 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, 4938–4949
24 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.* 88, 807–814
25 Oprea, T.J. and Gottfries, J. (2000) Towards minimalistic modelling of oral drug absorption. *J. Mol. Graph. Model.* 17, 261–274
26 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
27 Zhao, Y.H. *et al.* (2001) Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure-activity relationship (QSAR) with the Abraham descriptors. *J. Pharm. Sci.* 90, 749–784
28 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 Target.* 2, 151–165
29 Clark, D.E. (1999) Rapid calculation of polar surface area and its application to the prediction of transport phenomena. 2. Prediction of blood brain barrier penetration. *J. Pharm. Sci.* 88, 815–821
30 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
31 Kelder, J. *et al.* (1999) Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. *Pharm. Res.* 16, 1514–1519
32 Crivori, P. *et al.* (2000) Predicting blood-brain barrier permeation from three-dimensional molecular structure. *J. Med. Chem.* 43, 2204–2216
33 Platts, J.A. *et al.* (2001) Correlation and prediction of a large blood-brain distribution set – an LFER study. *Eur. J. Med. Chem.* 36, 719–730
34 Cronin, M.T.D. *et al.* (1999) Investigation of the mechanism of flux across human skin *in vitro* in quantitative structure-permeability relationships. *Eur. J. Pharm. Sci.* 7, 325–330
35 Potts, R.O. and Guy, R.H. (1995) A predictive algorithm for skin permeability: The effects of molecular size and hydrogen bond activity. *Pharm. Res.* 12, 1628–1633
36 Pugh, W.J. *et al.* (2000) Epidermal permeability penetrant structure relationships: 4, QSAR of permeant diffusion across human stratum corneum in terms of molecular weight, H-bonding and electronic charge. *Int. J. Pharm.* 197, 203–211
37 Abraham, M.H. *et al.* (1997) Algorithms for skin permeability using hydrogen bond descriptors: the problem of steroids. *J. Pharm. Pharmacol.* 49, 858–865
38 Prausnitz, M.R. and Noonan, J.S. (1998) Permeability of cornea, sclera, and conjunctiva: a literature analysis for drug delivery to the eye. *J. Pharm. Sci.* 87, 1479–1488
39 Raevsky, O.A. (1997) Hydrogen Bond Strength Estimation by means of HYBOT. In *Computer-Assisted Lead Finding and Optimisation* (Waterbeemd, H. *et al.*, eds.), pp. 367–378, Verlag, Basel
40 Abraham, M.H. and Platts, J.A. (2001) Hydrogen bond structural group constants. *J. Org. Chem.* 66, 3484–3491
41 Lombardo, F. *et al.* (2000) ElogP_{oct}: a tool for lipophilicity determination in drug discovery. *J. Med. Chem.* 43, 2922–2928
42 Du, C.M. *et al.* (2000) Characterising the selectivity of stationary phases and organic modifiers in reversed-phase high-performance liquid chromatographic systems by a general solvation equation using gradient elution. *J. Chromatogr. Sci.* 38, 503–511
43 Zissimos, A.M. *et al.* A comparison between the two general sets of linear free energy descriptors of Abraham and Klamt. *J. Chem. Inf. Comput. Sci.* (in press)
44 Zissimos, A.M. *et al.* (2002) Calculation of Abraham descriptors from solvent-water partition coefficients in four different systems; evaluation of different methods of calculation. *J. Chem. Soc. Perkin Trans.* 2, 470–477
45 Platts, J.A. *et al.* (1999) Estimation of molecular linear free energy relation descriptors by a group contribution approach. *J. Chem. Inf. Comput. Sci.* 39, 835–845
46 Abraham, M.H. *et al.* (1994) Hydrogen bonding. 33. Factors that influence the distribution of solutes between blood and brain. *J. Pharm. Sci.* 83, 1450–1456
47 Ertl, P. *et al.* (2000) Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* 43, 3714–3717
48 Gratton, J.A. *et al.* (1997) Molecular factors influencing drug transfer across the blood brain barrier. *J. Pharm. Pharmacol.* 49, 1211–1216
49 Abraham, M.H. and Le, J. (1999) The correlation and prediction of the solubility of compounds in water using an amended solvation energy relationship. *J. Pharm. Sci.* 88, 868–880
50 Abraham, M.H. *et al.* (2002) On the mechanism of human intestinal absorption. *Eur. J. Med. Chem.* 37, 595–605
51 Veber, D.F. *et al.* (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* 45, 2615–2623
52 Japertas, P. *et al.* (2002) Fragmental methods in the design of new compounds. Applications of the algorithm builder. *Quant. Struct. Act. Relation.* 21, 23–37