ELSEVIER

Advanced Drug Delivery Reviews 46 (2001) 3–26

advanced drug delivery reviews

www.elsevier.com/locate/drugdeliv

Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings☆

Christopher A. Lipinski*, Franco Lombardo, Beryl W. Dominy, Paul J. Feeney

Central Research Division, Pfizer Inc., Groton, CT 06340, USA

Received 9 August 1996; accepted 14 August 1996

Abstract

Experimental and computational approaches to estimate solubility and permeability in discovery and development settings are described. In the discovery setting 'the rule of 5' predicts that poor absorption or permeation is more likely when there are more than 5 H-bond donors, 10 H-bond acceptors, the molecular weight (MWT) is greater than 500 and the calculated Log P (CLogP) is greater than 5 (or MlogP > 4.15). Computational methodology for the rule-based Moriguchi Log P (MLogP) calculation is described. Turbidimetric solubility measurement is described and applied to known drugs. High throughput screening (HTS) leads tend to have higher MWT and Log P and lower turbidimetric solubility than leads in the pre-HTS era. In the development setting, solubility calculations focus on exact value prediction and are difficult because of polymorphism. Recent work on linear free energy relationships and Log P approaches are critically reviewed. Useful predictions are possible in closely related analog series when coupled with experimental thermodynamic solubility measurements.

Keywords: Rule of 5; Computational alert; Poor absorption or permeation; MWT; MLogP; H-Bond donors and acceptors; Turbidimetric solubility; Thermodynamic solubility; Solubility calculation

Contents

☆PII of original article: S0169-409X(96)00423-1. The article was originally published in *Advanced Drug Delivery Reviews* 23 (1997) 3–25.

*Corresponding author. Tel.: +1-860-4413561.

E-mail address: LIPINSKI@PFIZER.COM (C.A. Lipinski).

0169-409X/01/$ – see front matter
PII: S0169-409X(00)00129-0

1. Introduction

This review presents distinctly different but complementary experimental and computational approaches to estimate solubility and permeability in drug discovery and drug development settings. In the discovery setting, we describe an experimental approach to turbidimetric solubility measurement as well as computational approaches to absorption and permeability. The absence of discovery experimental approaches to permeation measurements reflects the authors' experience at Pfizer Central Research. Accordingly, the balance of poor solubility and poor permeation as a cause of absorption problems may be significantly different at other drug discovery locations, especially if chemistry focuses on peptidic-like compounds. This review deals only with solubility and permeability as barriers to absorption. Intestinal wall active transporters and intestinal wall metabolic events that influence the measurement of drug bioavailability are beyond the scope of this review. We hope to spark lively debate with our hypothesis that changes in recent years in medicinal chemistry physical property profiles may be the result of leads generated through high throughput screening. In the development setting, computational approaches to estimate solubility are critically reviewed based on current computational solubility research and experimental solubility measurements.

2. The drug discovery setting

2.1. Changes in drug leads and physico-chemical properties

In recent years, the sources of drug leads in the pharmaceutical industry have changed significantly. From about 1970 on, what were considered at that time to be large empirically-based screening programs became less and less important in the drug industry as the knowledge base grew for rational drug design [1]. Leads in this era were discovered using both in vitro and primary in vivo screening assays and came from sources other than massive primary in vitro screens. Lead sources were varied coming from natural products; clinical observations of drug side effects [1]; published unexamined patents; presentations and posters at scientific meetings; published reports in scientific journals and collaborations with academic investigators. Most of these lead sources had the common theme that the 'chemical lead' already had undergone considerable scientific investigation prior to being identified as a drug lead. From a physical property viewpoint, the most poorly behaved compounds in an analogue series were eliminated and most often the starting lead was in a range of physical properties consistent with the previous historical record of discovering orally active compounds.

This situation changed dramatically about 1989–1991. Prior to 1989, it was technically unfeasible to screen for in vitro activity across hundreds of thousands of compounds, the volume of random screening required to efficiently discover new leads. With the advent of high throughput screening in the 1989–1991 time period, it became technically feasible to screen hundreds of thousands of compounds across in vitro assays [2–4]. Combinatorial chemistry soon began[1] and allowed automated synthesis of massive numbers of compounds for screening in the new HTS screens. The process was accelerated by the rapid progress in molecular genetics which made possible the expression of animal and human receptor subtypes in cells lacking receptors that might interfere with an assay and by the construction of receptor constructs to facilitate signal detection. The screening of very large numbers of compounds necessitated a radical departure from the traditional method of drug solubilization. Compounds were no longer solubilized in aqueous media under thermodynamic equilibrating conditions. Rather, compounds were dissolved in dimethyl sulfoxide (DMSO) as stock solutions, typically at about 20–30 mmol and then were serially diluted into 96-well plates for assays (perhaps with some non ionic surfactant to improve solubility). In this paradigm, even very insoluble drugs could be tested because the kinetics of compound crystallization determined the apparent 'solubility' level. Moreover, compounds could partition into assay components such as membrane particulate material or cells or could bind to protein attached to the walls of the wells in the assay plate. The net effect was a screening technology for compounds in the μM concentration range that was largely divorced from the compounds true aqueous thermodynamic solubility. The apparent 'solubility' in the HTS screen is always higher, sometimes dramatically so, than the true thermodynamic solubility achieved by equilibration of a well characterized solid with aqueous media. The in vitro HTS testing process is quite reproducible and potential problems related to poor compound solubility are often compensated for by the follow-up to the primary screen. This is typically a more careful, more labor-intensive process of in vitro retesting to determine IC50s from dose response curves with more attention paid to solubilization. The net result of all these testing changes is that in vitro activity is reliably detected in compounds with very poor thermodynamic solubility properties. A corollary result is that the measurement of the true thermodynamic aqueous solubility is not very relevant to the screening manner in which leads are detected.

[1]A search through SciSearch and Chemical Abstracts for references to combinatorial chemistry in titles or descriptors using the truncated terms COMBIN? and CHEMISTR? gave the following number of references respectively: 1990, 0 and 0; 1991, 2 and 1; 1993, 8 and 8; 1994, 12 and 11; 1995, 46 and 45.

2.2. Factors affecting physico-chemical lead profiles

The physico-chemical profile of current leads i.e. the 'hits' in HTS screens now no longer depends on compound solubility sufficient for in vivo activity but depends on: (1) the medicinal chemistry principles relating structure to in vitro activity; (2) the nature of the HTS screen; (3) the physico-chemical profile of the compound set being screened and (4) to human decision making, both overt and hidden as to the acceptability of compounds as starting points for medicinal chemistry structure activity relationship (SAR) studies.

One of the most reliable methods in medicinal chemistry to improve in vitro activity is to incorporate properly positioned lipophilic groups. For example, addition of a single methyl group that can occupy a receptor 'pocket' improves binding by about 0.7 kcal/mol [6]. By way of contrast, it is generally difficult to improve in vitro potency by manipulation of the polar groups that are involved in ionic receptor interactions. The interaction of a polar group in a drug with solvent versus interaction with the target receptor is a 'wash' unless positioning of the polar group in the drug is precise. The traditional lore is that the lead has the polar groups in the correct (or almost correct) position and that in vitro potency is improved by correctly positioned lipophilic groups that occupy receptor pockets. Polar groups in the drug that are not required for binding can be tolerated if they occupy solvent space but they do not add to receptor binding. The net effect of these simple medicinal chemistry principles is that, other factors being equal, compounds with correctly

positioned polar functionality will be more readily detectable in HTS screens if they are larger and more lipophilic.

The nature of the screen determines the physico-chemical profile of the resultant 'hits'. The larger the number of hits that are detected, the more the physico-chemical profile of the 'hits' resembles the overall compound set being screened. Technical factors such as the design of the screen and human cultural factors such as the stringency of the evaluation as to what is a suitable lead worth are major determinants of the physico-chemical profiles of the eventual leads. Screens designed with very high specificity, for example many receptor based assays, generate small numbers of hits in the μM range. In these types of screens the signal is easy to detect against background noise, the hits are few or can be made few by altering potency criteria and the physico-chemical profiles tend towards more lipophilic, larger, less soluble compounds. Tight control of the criteria for activity detection in the initial HTS screen minimizes labor-intensive secondary evaluation and minimizes the effect of human biases. The downside is that lower potency hits with more favorable physico-chemical property profiles may be discarded.

Cell-based assays, by their very nature tend to produce more 'hits' than receptor-based screens. These types of assays monitor a functional event, for example a change in the level of a signaling intermediate or the expression level of M-RNA or protein. Multiple mechanisms may lead to the measured end point and only a few of these mechanisms may be desirable. This leads to a larger number of hits and therefore their physico-chemical profile will more closely resemble that of the compound set being screened. Perhaps, equally importantly, a larger volume of secondary evaluation allows for a greater expression of human bias. Bias is especially difficult to quantify in the chemists perception of a desirable lead structure.

The physico-chemical profile of the compound set being screened is the first filter in the physico-chemical profile of an HTS 'hit'. Obviously high molecular weight, high lipophilicity compounds will not be detected by a screen if they are not present in the library. In the real world, trade-offs occur in the choice of profiles for compound sets. An exclusively low molecular weight, low lipophilicity library likely increases the difficulty of detecting 'hits' but simplifies the process of discovering an orally active drug once the lead is identified. The converse is true of a high molecular weight high lipophilicity library. In our experience, commercially available (non combinatorial) compounds like those available from chemical supply houses tend towards lower molecular weights and lipophilicities.

Human decision making, both overt and hidden can play a large part in the profile of HTS 'hits'. For example, a requirement that 'hits' possess an acceptable range of measured or calculated physico-chemical properties will obviously affect the starting compound profiles for medicinal chemistry SAR. Less obvious are hidden biases. Are the criteria for a 'hit' changing to higher potency (lower IC50) as the HTS screen runs? Labor-intensive secondary follow-up is decreased but less potent, perhaps physico-chemically more attractive leads, may be eliminated. How do chemists react to potential lead structures? In an interesting experiment, we presented a panel of our most experienced medicinal chemists with a group of theoretical lead structures — all containing literature 'toxic' moieties. Our chemists split into two very divergent groups; those who saw the toxic moieties as a bar to lead pursuit and those who recognized the toxic moiety but thought they might be able to replace the offending moiety. An easy way to illustrate the complexity of the chemists perception of lead attractiveness is to examine the remarkably diverse structures of the new chemical entities (NCEs) introduced to market that appear at the back of recent volumes of *Annual Reports in Medicinal Chemistry*. No single pharmaceutical company can conduct research in all therapeutic areas and so some of these compounds, which are all marketed drugs, will inevitably be less familiar and potentially less desirable to the medicinal chemist at one research location, but may be familiar and desirable to a chemist at another research site.

2.3. Identifying a library with favorable physico-chemical properties

The idea in selecting a library with good absorption properties is to use the clinical Phase II selection process as a filter. Drug development is expensive

and the most poorly behaved compounds are weeded out early. Our hypothesis was that poorer physico-chemical properties would predominate in the many compounds that enter into and fail to survive pre-clinical stages and Phase I safety evaluation. We expected that the most insoluble and poorly permeable compounds would have been eliminated in those compounds that survived to enter Phase II efficacy studies. We could use the presence of United States Adopted Name (USAN) or International Non-proprietary Name (INN) names to identify compounds entering Phase II since most drug companies (including Pfizer) apply for these names at entry to Phase II.

The (WDI) World Drug Index is a very large computerized database of about 50 000 drugs from the Derwent Co. The process used to select a subset of 2245 compounds from this database that are likely to have superior physico-chemical properties is as follows: From the 50 427 compounds in the WDI File, 7894 with a data field for a USAN name were selected as were 6320 with a data field for an INN. From the two lists, 8548 compounds had one or both USAN or INN names. These were searched for a data field 'indications and usage' suggesting clinical exposure, resulting in 3704 entries. From the 3704 using a substructure data field we eliminated 1176 compounds with the text string 'POLY', 87 with the text string 'PEPTIDE' and 101 with the text string 'QUAT'. Also eliminated were 53 compounds containing the fragment O = P-O. We coined the term 'USAN' library for this collection of drugs.

2.4. The target audience — medicinal chemists

Having identified a library of drugs selected by the economics of entry to the Phase II process we sought to identify calculable parameters for that library that were likely related to absorption or permeability. Our approach and choice of parameters was dictated by very pragmatic considerations. We wanted to set up an absorption–permeability alert procedure to guide our medicinal chemists. Keeping in mind our target audience of organic chemists we wanted to focus on the chemists very strong pattern recognition and chemical structure recognition skills. If our target audience had been pharmaceutical scientists we would not have deliberately excluded equations or regression coefficients. Experience had taught us that a focus on the chemists very strong skills in pattern recognition and their outstanding chemistry structural recognition skills was likely to enhance information transfer. In effect, we deliberately emphasized enhanced educational effectiveness towards a well defined target audience at the expense of a loss of detail. Tailoring the message to the audience is a basic communications principle. One has only to look at the popular chemistry abstracting booklets with their page after page of chemistry structures and minimal text to appreciate the chemists structural recognition skills. We believe that our chemists have accepted our calculations at least in part because the calculated parameters are very readily visualized structurally and are presented in a pattern recognition format.

2.5. Calculated properties of the 'USAN' library

Molecular weight (formula weight in the case of a salt) is an obvious choice because of the literature relating poorer intestinal and blood brain barrier permeability to increasing molecular weight [7,8] and the more rapid decline in permeation time as a function of molecular weight in lipid bi-layers as opposed to aqueous media [9]. The molecular weights of compounds in the 2245 USANs were lower than those in the whole 50 427 WDI data set. In the USAN set 11% had MWTs $>$ 500 compared to 22% in the entire data set. Compounds with MWT $>$ 600 were present at 8% in the USAN set compared to 14% in the entire data set. This difference is not explainable by the elimination of the very high MWTs in the USAN selection process. Rather it reflects the fact that higher MWT compounds are in general less likely to be orally active than lower MWTs.

Lipophilicity expressed as a ratio of octanol solubility to aqueous solubility appears in some form in almost every analysis of physico-chemical properties related to absorption [10]. The computational problem is that an operationally useful computational alert to possible absorption–permeability problems must have a no fail log P calculation. In our experience, the widely used and accurate Pomona College Medicinal Chemistry program applied to our compound file failed to provide a calculated log P (CLogP) value because of missing fragments for at

least 25% of compounds. The problem is not an inordinate number of 'strange fragments' in our chemistry libraries but rather lies in the direction of the trade off between accuracy and ability to calculate all compounds adopted by the Pomona College team. The CLogP calculation emphasizes high accuracy over breadth of calculation coverage. The fragmental CLogP value is defined with reference to five types of intervening isolating carbons between the polar fragments. As common a polar fragment as a sulfide (-S-) linkage generates missing fragments when flanked by rare combinations of the isolating carbon types. Polar fragments as defined by the CLogP calculation can be very large and are not calculated as the sum of smaller, more common, polar fragments. This approach enhances accuracy but increases the number of missing fragments.

We implemented the log P calculation (MLogP) as described by Moriguchi et al. [11] within the Molecular Design Limited MACCS and ISIS base programs to avoid the missing fragment problem. As a rule-based system, the Moriguchi calculation always gives an answer. The pros and cons of the Moriguchi algorithm have been debated in the literature [12,13]. We recommend that, within analog series, our medicinal chemists use the more accurate Pomona CLogP calculation if possible. For calculation or tracking of library properties the less accurate MLogP program is used.

Only about 10% of USAN compounds have a CLogP over 5. The CLogP value of 5 calculated on the USAN data set corresponds to an MLogP of 4.15. The slope of CLogP (*x* axis) versus MLogP (*y* axis) is less than unity. At the high log P end, the Moriguchi MLogP is somewhat lower than the MedChem CLogP. In the middle log P range at about 2, the two scales are similar. Experimentally there is almost certainly a lower (hydrophilic) log P limit to absorption and permeation. Operationally, we have ignored a lower limit because of the errors in the MLogP calculation and because excessively hydrophilic compounds are not a problem in compounds originating in our medicinal chemistry laboratories.

An excessive number of hydrogen bond donor groups impairs permeability across a membrane bi-layer [14,15]. Hydrogen donor ability can be measured indirectly by the partition coefficient between strongly hydrogen bonding solvents like water or ethylene glycol and a non hydrogen bond accepting solvent like a hydrocarbon [15] or as the log of the ratio of octanol to hydrocarbon partitioning. In vitro systems for studying intestinal drug absorption have been recently reviewed [16]. Computationally, hydrogen donor ability differences can be expressed by the solvatochromic α parameter of a donor group with perhaps a steric modifier to allow for the interactions between donor and acceptor moieties. Experimental α values for hydrogen bond donors and β values for acceptor groups [17] have been compiled by Professor Abraham in the UK and by the Raevsky group in Russia [18,19]. Both research groups currently express the hydrogen bond donor and acceptor properties of a moiety on a thermodynamic free energy scale. In the Raevsky C scale, donors range from about -4.0 for a very strong donor to -0.5 for a very weak donor. Acceptors values in the Raevsky C scale are all positive and range from about 4.0 for a strong acceptor to about 0.5 for a weak acceptor. In the Abraham scale both donors and acceptors have positive values that are about one-quarter of the absolute C values in the Raevsky scale.

We found that simply adding the number of NH bonds and OH bonds does remarkably well as an index of H bond donor character. Importantly, this parameter has direct structural relevance to the chemist. When one looks at the USAN library there is a sharp cutoff in the number of compounds containing more than 5 OHs and NHs. Only 8% have more than 5. So 92% of compounds have five or fewer H bond donors and it is the smaller number of donors that the literature links with better permeability.

Too many hydrogen bond acceptor groups also hinder permeability across a membrane bi-layer. The sum of Ns and Os is a rough measure of H bond accepting ability. This very simple calculation is not nearly as good as the OH and NH count (as a model for donor ability) because there is far more variation in hydrogen bond acceptor than donor ability across atom types. For example, a pyrrole and pyridine nitrogen count equally as acceptors in the simple N O sum calculation even though a pyridine nitrogen is a very good acceptor (2.72 on the C scale) and the pyrrole nitrogen is an far poorer acceptor (1.33 on the C scale). The more accurate solvatochromic β parameter which measures acceptor ability varies far

more on a per nitrogen or oxygen atom basis than the corresponding α parameter. When we examined the USAN library we found a fairly sharp cutoff in profiles with only about 12% of compounds having more than 10 Ns and Os.

2.6. The 'rule of 5' and its implementation

At this point we had four parameters that we thought should be globally associated with solubility and permeability; namely molecular weight; Log P; the number of H-bond donors and the number of H-bond acceptors. In a manner similar to setting the confidence level of an assay at 90 or 95% we asked how these four parameters needed to be set so that about 90% of the USAN compounds had parameters in a calculated range associated with better solubility or permeability. This analysis led to a simple mnemonic which we called the 'rule of 5' [20] because the cutoffs for each of the four parameters were all close to 5 or a multiple of 5. In the USAN set we found that the sum of Ns and Os in the molecular formula was greater than 10 in 12% of the compounds. Eleven percent of compounds had a MWT of over 500. Ten percent of compounds had a CLogP larger than 5 (or an MLogP larger than 4.15) and in 8% of compounds the sum of OHs and NHs in the chemical structure was larger than 5. The 'rule of 5' states that: poor absorption or permeation are more likely when:

There are more than 5 H-bond donors (expressed as the sum of OHs and NHs);
The MWT is over 500;
The Log P is over 5 (or MLogP is over 4.15);
There are more than 10 H-bond acceptors (expressed as the sum of Ns and Os);
Compound classes that are substrates for biological transporters are exceptions to the rule.

When we examined combinations of any two of the four parameters in the USAN data set, we found that combinations of two parameters outside the desirable range did not exceed 10%. The exact values from the USAN set are: sum of N and O + sum of NH and OH — 10%; sum of N and O + MWT — 7%; sum of NH and OH + MWT — 4% and sum of MWT + Log P — 1%. The rarity (1%) among USAN drugs of the combination of high MWT and high log P was striking because this particular combination of physico-chemical properties in the USAN list is enhanced in the leads resulting from high throughput screening.

The rule of 5 is now implemented in our registration system for new compounds synthesized in our medicinal chemistry laboratories and the calculation program runs automatically as the chemist registers a new compound. If two parameters are out of range, a 'poor absorption or permeability is possible' alert appears on the registration screen. All new compounds are registered and so the alert is a very visible educational tool for the chemist and serves as a tracking tool for the research organization. No chemist is prevented from registering a compound because of the alert calculation.

2.7. Orally active drugs outside the 'rule of 5' mnemonic and biologic transporters

The 'rule of 5' is based on a distribution of calculated properties among several thousand drugs. Therefore by definition, some drugs will lie outside the parameter cutoffs in the rule. Interestingly, only a small number of therapeutic categories account for most of the USAN drugs with properties falling outside our parameter cutoffs. These orally active therapeutic classes outside the 'rule of 5' are: antibiotics, antifungals, vitamins and cardiac glycosides. We suggest that these few therapeutic classes contain orally active drugs that violate the 'rule of 5' because members of these classes have structural features that allow the drugs to act as substrates for naturally occurring transporters. When the 'rule of 5' is modified to exclude these few drug categories only a very few exceptions can be found. For example, among the NCEs between 1990 and 1993 falling outside the double cutoffs in 'the rule of 5', there were nine non-orally active drugs and the only orally active compounds outside the double cutoffs were seven antibiotics. Fungicides–protoazocides–antiseptics also fall outside the rule. For example, among the 41 USAN drugs with MWT $>$ 500 and MLogP $>$ 4.15 there were nine drugs in this class. Vitamins are another orally active class drug with parameter values outside the double cutoffs. Close to 100

vitamins fell into this category. Cardiac glycosides, an orally active drug class also fall outside the parameter limits of the rule of 5. For example among 90 USANs with high MWT and low MLogP there were two cardiac glycosides.

2.8. High MWT USANs and the trend in MLogP

In our USAN data set we plotted MLogP against MWT and examined the compound distributions as defined by the 50 and 90% probability ellipses. A large number of USAN compounds had MLogP more negative than -0.5. Among the USAN compounds there was a trend for higher MWT to correlate with lower MLogP. This type of trend is distinctly different from the positive correlation between MLogP and MWT found in most SAR data sets. Usually as MWT increases, compound lipophilicity increases and MLogP becomes larger (more positive). From among the 2641 USANs, we selected the 405 with MLogP more negative than -0.5 and from among these selected those with MWT in excess of 500 and mapped the resulting 90 against therapeutic activity fields in the MACCS WDI database. About one half (44 of 90) of these high MWT, low MLogP USANs were orally inactive consisting of 26 peptide agonists or antagonists, 11 quaternary ammonium salts and seven miscellaneous non-orally active agents.

Among the USAN compounds in our list fewer than 10% of compounds had either high MLogP or high MWT. The combination of both these properties in the same compound was even rarer. Among 2641 USANs there were only 41 drugs with MWT $>$ 500 and MLogP $>$ 4.15, about one-half (21) were orally inactive. Among the remainder there were only six orally active compounds not in the fungicide and vitamin classes.

2.9. New chemical entities, calculations

New chemical entities introduced between 1990 and 1993 were identified from a summary listing in vol. 29 of *Annual Reports in Medicinal Chemistry*. All our computer programs for calculating physicochemical properties require that the compound be described in computer-readable format. We mapped compound names and used structural searches to identify 133 of the NCEs in the Derwent World Drug to give us the computer-readable formats to calculate the rule of 5. The means of calculated properties were well within the acceptable range. The average Moriguchi log P was 1.80, the sum of H-bond donors was 2.53, the molecular weight was 408 and the sum of Ns and Os was 6.95. The incidence of alerts for possible poor absorption or permeation was 12%.

2.10. Drugs in absorption and permeability studies, calculations

Very biased data sets are encountered in the types of drugs that are reported in the absorption or permeability literature. Calculated properties are quite favorable when compared to the profiles of compounds detected by high throughput screening. Compounds that are studied are usually orally active marketed drugs and therefore by definition have properties within the acceptable range. What is generally not appreciated is that absorption and permeability are mostly reported for the older drugs. For example, our list of compounds with published literature on absorption or permeability, studied internally for validation purposes, is highly biased against NCEs. Only one drug in our list of 73 was introduced in the period 1990 to date. In part this reflects drug availability, since drugs under patent are not sold by third parties. Drugs studied in absorption or permeability models tend to be those with value for assay validation purposes, i.e. those with considerable pre-existing literature. In addition, some of the newer studies are driven by a regulatory agency interest in the permeability properties of generic drugs. In our listing of 73 drugs in absorption or permeability studies there are 33 generic drugs whose properties the FDA is currently profiling. Our list includes an additional 23 drugs with CACO-2 cell permeation data. Most of these are from the speakers' handouts at a recent meeting on permeation prediction [21]; a few are from internal Pfizer CACO-2 studies. A final 12 drugs are those with zwitterionic or very hydrophilic properties for which there are either literature citations or internal Pfizer data. The means of calculated properties for compounds in this list are well within the acceptable range. The average Moriguchi log P was 1.60, the sum of H-bond donors was 2.49, the molecular

weight was 361 and the sum of Ns and Os was 6.27. The incidence of alerts for possible poor absorption or permeation was 12% (Table 1).

2.11. Validating the computational alert

Validating a computational alert for poor absorption or permeation in a discovery setting is quite different than validating a quantitative prediction calculation in a developmental setting. In effect, a discovery alert is a very coarse filter that identifies compounds lying in a region of property space where the probability of useful oral activity is very low. The goal is to move chemistry SAR towards the region of property space where oral activity is reasonably possible (but not assured) and where the more labor-intensive techniques of drug metabolism and the pharmaceutical sciences can be more efficiently employed. A compound that fails the computational alert will likely be poorly bio-available because of poor absorption or permeation and lies within that region of property space where good absorption or solubility is unlikely. We believe the alert has its primary value in identifying problem compounds. In our experience, most compounds failing the alert also will prove troublesome if they progress far enough to be studied experimentally. However, the converse is not true. Compounds passing the alert still can prove troublesome in experimental studies.

In this perspective, a useful computational alert correctly identifies drug projects with known absorption problems. Drugs in human therapy, whether poorly or well absorbed from the viewpoint of the pharmaceutical scientist, should profile as ‘drugs’, i.e. as having reasonable prospects for oral activity. The larger the computational and experimental difference between drugs in human therapy and those which are currently being made in medicinal chemistry laboratories, the greater the confidence that the differences are meaningful. We assert that absorption problems have recently become worse in the pharmaceutical industry as attested to by recent meetings and symposia on this subject [22] and by the informal but industry-wide concern of pharmaceutical scientists about drug candidates with less than optimal physical properties. If we are correct, within any drug organization, one should be able to quantify by calculation whether time-dependent changes that might impair absorption have occurred in medicinal chemistry. If these changes have occurred one can try to correlate these with changes in screening strategy.

2.12. Changes in calculated physical property profiles at Pfizer

How relevant is our experience at the Pfizer Central Research laboratories in Groton to what may be expected to be observed in other drug discovery organizations? The physical property profiles of drug leads discovered through HTS will be similar industry-wide to the extent that testing methodology, selection criteria and the compounds being screened are similar. Changes in physical property profiles of synthetic compounds, made in follow-up of HTS leads by medicinal laboratories, depend on the timing of a major change towards HTS screening. The Pfizer laboratories in Groton were one of the first to realize and implement the benefits of HTS in lead detection. As a consequence, we also have been one of the first to deal with the effects of this change in screening strategy on physico-chemical properties. In Groton, 1989 marked the beginning of a significant change towards HTS screening. This process was largely completed by 1992 and currently HTS is now the major, rich source of drug discovery leads and has largely supplanted the pre-1989 pattern of lead generation.

At the Pfizer Groton site, we have retrospectively examined the MWT distributions of compounds made in the pre-1989 era and since 1989. Since our registration systems unambiguously identify the source of each compound, we can identify any time-dependent change in physical properties and we can compare the profiles of internally synthesized compounds with the profiles of compounds purchased from external commercial sources.

Before 1989, the percentage of internally synthesized high MWT compounds oscillated in a range very similar to the USAN library (Table 2). Starting in 1989, there was an upward jump in the percentage of high MWT compounds and a further jump in 1992 to a new stable MWT plateau that is higher than in the USAN library and higher than any yearly oscillation in the pre-1989 era. By contrast, there was no

Table 1
Partial list of drugs in absorption and permeability studies

Drug name	MLogP	OH+NH[c]	MWT	N+O[d]	Alert[e]
Aciclovir[a,b]	−0.09	4	225.21	8	0
Alprazolam[a]	4.74	0	308.77	4	0
Aspirin[b]	1.70	1	180.16	4	0
Atenolol[a,b]	0.92	4	266.34	5	0
Azithromycin[b]	0.14	5	749.00	14	1
AZT[a]	−4.38	2	267.25	9	0
Benzyl-penicillin[b]	1.82	2	334.40	6	0
Caffeine[b]	0.20	0	194.19	6	0
Candoxatril[b]	3.03	2	515.65	8	0
Captopril[a]	0.64	1	217.29	4	0
Carbamazepine[a]	3.53	2	236.28	3	0
Chloramphenicol[b]	1.23	3	323.14	7	0
Cimetidine[a,b]	0.82	3	252.34	6	0
Clonidine[b]	3.47	2	230.10	3	0
Cyclosporine[a]	−0.32	5	1202.64	23	1
Desipramine[a,b]	3.64	1	266.39	2	0
Dexamethasone[b]	1.85	3	392.47	5	0
Diazepam[b]	3.36	0	284.75	3	0
Diclofenac[a]	3.99	2	296.15	3	0
Diltiazem-HCl[a]	2.67	0	414.53	6	0
Doxorubicin[b]	−1.33	7	543.53	12	1
Enalapril-maleate[a]	1.64	2	376.46	7	0
Erythromycin[b]	−0.14	5	733.95	14	1
Famotidine[a]	−0.18	8	337.45	9	0
Felodipine[a,b]	3.22	1	384.26	5	0
Fluorouracil[b]	−0.63	2	130.08	4	0
Flurbiprofen[a]	3.90	1	244.27	2	0
Furosemide[a]	0.95	4	330.75	7	0
Glycine[b]	−3.44	3	75.07	3	0
Hydrochlorthiazide[a]	−1.08	4	297.74	7	0
Ibuprofen[b]	3.23	1	206.29	2	0
Imipramine[b]	3.88	0	280.42	2	0
Itraconazole[a]	5.53	0	705.65	12	1
Ketaconazole[a]	4.45	0	380.92	1	0
Ketoprofen[a]	3.37	1	254.29	3	0
Labetalol-HCl[a]	2.67	5	328.42	5	0
Lisinopril[a]	1.11	5	405.50	8	0
Mannitol[b]	−2.50	6	182.18	6	0
Methotrexate[b]	1.60	7	454.45	13	1
Metoprolol-tartrate[a,b]	1.65	2	267.37	4	0
Nadolol[a]	0.97	4	309.41	5	0
Naloxone[b]	1.53	2	327.38	5	0
Naproxen-sodium[a,b]	2.76	1	230.27	3	0
Nortriptylene-HCl[a]	4.14	1	263.39	1	0
Omeprazole[a]	−4.38	2	267.25	9	0
Phenytoin[a]	2.20	2	451.49	10	0
Piroxicam[a]	0.00	2	331.35	7	0
Prazosin[b]	2.05	2	383.41	9	0
Propranolol-HCl[a,b]	2.53	2	259.35	3	0
Quinidine[b]	2.19	1	324.43	4	0
Ranitidine-HCl[a]	0.66	2	314.41	7	0
Scopolamine[b]	1.42	1	303.36	5	0
Tenidap[b]	1.95	2	320.76	5	0
Terfenadine[a]	4.94	2	471.69	3	0
Testosterone[b]	3.70	1	288.43	2	0
Trovafloxacin[b]	2.81	3	416.36	7	0
Valproic-acid[b]	2.06	1	144.22	2	0
Vinblastine[b]	2.96	3	811.00	13	1
Ziprasidone[b]	3.71	1	412.95	5	0

[a]Standard or drug in FDA bioequivalence study.

[b]Studied in CACO-2 permeation.

[c]Sum of OH and NH H-bond donors.

[d]Sum of N and O H-bond acceptors.

[e]Computational alert according to the rule of 5; 0, no problem detected; 1, poor absorption or permeation are more likely.

Table 2
Percent of compounds with MWT (including salt) above 500

Year registered	Synthetic compounds	Commercial compounds
Pre-1984	16.0	5.4
1984	18.9	14.7
1985	12.1	15.5
1986	12.6	5.5
1987	13.4	5.8
1988	14.6	8.2
1989	23.4	4.1
1990	21.1	3.3
1991	25.4	1.8
1992	34.2	6.8
1993	33.2	8.4
1994	32.7	7.9

change in the MWT profiles of commercially purchased compounds over the same time period. A comparison of the MWT and MLogP percentiles of synthetic compounds for a year before the advent of HTS and for 1994 in the post-HTS era shows a similar pattern (Table 3). The upper range percentiles for MWT and MLogP properties are skewed towards physical properties less favorable for oral absorption in the more recent time period.

The trend towards higher MWT and LogP is in the direction of the property mix that is least populated in the USAN library. There was no change over time in the population of compounds with high numbers of H-bond donors or acceptors.

2.13. The rationale for measuring drug solubility in a discovery setting

In recent years, we have been exploring experimental protocols in a discovery setting that measure drug solubility in a manner as close as possible to the actual solubilization process used in our biological laboratories. The rationale is that the physical forms of the compounds solubilized and the methods used to solubilize compounds in discovery are very different from those used by our pharmaceutical scientists and that mimicking the discovery process will lead to the best prediction of in vivo SAR.

Table 3
Synthetic compound properties in 1986 (pre-HTS) and 1994 (post-HTS)

Percentile	MLogP		MWT	
	1986	1994	1986	1994
90th	4.30	4.76	514	726
75th	3.48	3.90	415	535
50th	2.60	2.86	352	412

In discovery, the focus is on keeping a drug solubilized for an assay rather than on determining the solubility limit. Moreover, there is no known automated methodology that can efficiently solubilize hundreds of thousands of sometimes very poorly soluble compounds under thermodynamic conditions. In our biological laboratories, compounds that are not obviously soluble in water or by pH adjustment are pre-dissolved in a water miscible solvent (most often DMSO) and then added to a well stirred aqueous medium. The equivalent of a thermodynamic solubilization, i.e. equilibrating a solid compound for 24–48 h, separating the phases, measuring the soluble aqueous concentration and then using the aqueous in an assay, is not done. When compounds are diluted into aqueous media from a DMSO stock solution, the apparent solubility is largely kinetically driven. The influence of crystal lattice energy and the effect of polymorphic forms on solubility is, of course, completely lost in the DMSO dissolution process. Drug added in DMSO solution to an aqueous medium is delivered in a very high energy state which enhances the apparent solubility. The appearance of precipitate (if any) from a thermodynamically supersaturated solution is kinetically determined and to our knowledge is not predictable by computational methods. Solubility may also be perturbed from the true thermodynamic value in purely aqueous media by the presence of a low level of residual DMSO.

The physical form of the first experimental lot of a compound made in a medicinal chemistry lab can be very different from that seen by the pharmaceutical scientist at a later stage of development. Solution spectra, HPLC purity criteria and mass spectral analysis are quite adequate to support a structural assignment when the chemist's priority is on efficiently making as many well selected compounds as possible in sufficient quantity for in vitro and in vivo screening. All the measurements that support struc-

tural assignment are unaffected by the energy state (polymorphic form) of the solid. Indeed, depending on the therapeutic area, samples may not be crystalline and most compounds synthesized for the first time are unlikely to be in lower energy crystalline forms. Attempts to compute solubility using melting point information are not useful if samples do not have well defined melting points. Well characterized, low energy physical form (from a pharmaceutics viewpoint) reduces aqueous solubility and may actually be counter productive to the discovery chemists priority of detecting in vivo SAR.

In this setting, thermodynamic solubility data can be overly pessimistic and may mislead the chemist who is trying to relate chemical structural changes to absorption and oral activity in the primary in vivo assay. Our goal is to provide a relevant experimental solubility measurement so that chemistry can move from the pool of poorly soluble, orally inactive compounds towards those with some degree of oral activity. For maximum relevance to the in vivo biological assay our solubility measurement protocol is as close as possible to the biological assay 'solubilization'. In this paradigm, any problems that might be related to the poor absorption of a low energy crystalline solid under thermodynamic conditions are postponed and not solved. The efficiency gain in an early discovery stage solubility assay lies in the SAR direction provided to chemistry and in the more efficient application of drug metabolism and pharmaceutical sciences resources once oral activity is detected. The value of this type of assay is very stage-dependent and the discovery type of assay is not a replacement for a thermodynamic solubility measurement at a later stage in the discovery process.

2.14. Drugs have high turbidimetric solubility

Measuring solubility by turbidimetry violates almost every precept taught in the pharmaceutical sciences about 'proper' thermodynamic solubility measurement. Accordingly, we have been profiling known marketed drugs since our initial presentation on turbidimetric solubility measurement [23] and have measured turbidimetric solubilities on over 350 drugs from among those listed in the Derwent World Drug Index. The calculated properties of these drugs are well within the favorable range for oral absorption. The average of the calculated properties are: MLogP, 1.79; the sum of OH and NH, 2.01; MWT, 295.4; the sum of N and O, 4.69. Without regard to the therapeutic class, only 4% of these drugs would have been flagged as having an increased probability of poor absorption or permeability in our computational alert. Of the 353 drugs, 305 (87%) had a turbidimetric solubility of greater than 65 μg/ml. There were only 20 drugs (7%) with a turbidimetric solubility of 20 μg/ml or less. If turbidimetric solubility values lie in this low range, we suggest to our chemists that the probability of useful oral activity is very low unless the compound is unusually potent (e.g. projected clinical dose of 0.1 mg/kg) or unusually permeable (top tenth percentile in absorption rate constant) or unless the compound is a member of a drug class that is a substrate for a biological transporter.

Our drug list was compiled without regard to literature thermodynamic solubilities but does contain many of the types of compounds studied in the absorption literature. Of the 353 drugs studied in the discovery solubility assay, 171 are drugs from four sources. There are 77 drugs from the compilation of 200 drugs by Andrews et al. [6]. This compilation is biased towards drugs with reliable measured in vitro receptor affinity and with interesting functionality and not necessarily towards drugs with good absorption or permeation characteristics. There are 23 drugs from a list of generics whose properties FDA is currently profiling for bio-equivalency standards. In addition, there are 42 NCEs introduced between 1983 and 1993 and 37 entries are for drugs with CACO-2 cell permeation data.

The profile of drug turbidimetric solubilities serves as a useful benchmark. Compounds that are drugs have a very low computational alert rate for absorption or permeability problems and a low measured incidence of poor turbidimetric solubility of about 10%. The calculated profiles and alert rates of compounds made in medicinal chemistry laboratories can be compared to those of drugs and the profiles can be compared on a project by project basis.

Within the physical property manifold of 'marketed drugs' we would expect a poor correlation of our turbidimetric solubility data with literature thermodynamic solubility data since the properties of

'drugs' occupy only a small region of property space relative to what is possible in synthetic compounds and HTS 'hits'. Our turbidimetric solubilities for drugs are almost entirely at the top end of a relatively narrow solubility range, whereas from a thermodynamic viewpoint the drugs in our list cover a wide spectrum of solubility. We caution that turbidimetric solubility measurements are most definitely not a substitute for careful thermodynamic solubility measurements on well characterized crystalline drugs and should not be used for decision making in a development setting.

2.15. High throughput screening hits, calculations and solubility measurements

Calculated properties and measured turbidimetric solubilities for the best compounds identified as 'hits' in our HTS screens are in accord with the hypothesis that the physico-chemical profiles of leads have changes from those in the pre-1989 time period. Nearly 100 of the most potent 'hits' from our high throughput screens were examined computationally and their turbidimetric solubilities were measured. The profiles are strikingly different from those of the 353 drugs we studied. The HTS hits are on average more lipophilic and less soluble than the drugs. The 96 compounds we measured were the end product of detection in HTS screens and secondary in vitro evaluation. These were the compounds highlighted in summaries and which captured the chemist's interest with many IC50s clustered in the 1 μM range. As such, they are the product of a biological testing process and a chemistry evaluation as to interesting subject matter. Average MLogP for the HTS hits was a full log unit higher than for the drugs and the average MWT was nearly 50 Da higher. By contrast, there was little difference in the number of hydrogen bond donors and acceptors. The distribution curves for MLogP and MWT are roughly the same shape for the HTS hits and drugs but the means are shifted upwards in the HTS hits with a higher distribution of compounds towards the unfavorable range of physico-chemical properties. The actual averages, HTS vs. Drug are: MLogP, 2.81 vs. 1.79; MWT, 366 vs. 295; sum of OH NH, 1.80 vs. 2.01; sum of N and O, 5.4 vs. 4.69.

2.16. The triad of potency, solubility and permeability

Acceptable drug absorption depends on the triad of dose, solubility and permeability. Our computational alert does not factor in dose, i.e. drug potency. It only addresses properties that are related to potential solubility and permeation problems and it does not allow for a very favorable value of one parameter to compensate for a less favorable value of another parameter. In a successful marketed drug, one parameter can compensate for another. For example, a computational alert is calculated for azithromycin, a successful marketed antibiotic. In azithromycin, which has excellent oral activity, a very high aqueous solubility of 50 mg/ml more than counterbalances a very low absorption rate in the rat intestinal loop of 0.001 min^{-1}. Poorer permeability in orally active peptidic-like drugs is usually compensated by very high solubility. Our solubility guidelines to our chemists suggest a minimum thermodynamic solubility of 50 μg/ml for a compound that has a mid-range permeability and an average potency of 1.0 mg/kg. These solubility guidelines would be markedly higher if the average compound had low permeability.

2.17. Protocols for measuring drug solubility in a discovery setting

The method and timing of introduction of the drug into the aqueous media are key elements in our discovery solubility protocol. Drug is dissolved in DMSO at a concentration of 10 μg/μl of DMSO which is close to the 30 mM DMSO stock concentration used in our own biology laboratories. This is added a microlitre at a time to a non-chloride containing pH 7 phosphate buffer at room temperature. The decision to avoid the presence of chloride was a tradeoff between two opposing considerations. Biology laboratories with requirements for iso-osmotic media use vehicles containing physiological levels of saline (e.g. Dulbecco's phosphate buffered saline) with the indirect result that the solubility of HCl salts (by far the most frequent amine salt from our chemistry laboratories) can be depressed by the common ion effect. Counter to this consideration, is the near 100% success rate of our pharmaceutical

groups in replacing problematical HCl salts with other salts not subject to a chloride common ion effect. We chose the non-chloride containing medium to avoid pessimistic solubility values resulting from a historically very solvable problem.

The appearance of precipitate is kinetically driven and so we avoid a short time course experiment where we might miss precipitation that occurs on the type of time scale that would affect a biological experiment. The additions of DMSO are spaced a minute apart. A total of 14 additions are made. These correspond to solubility increments of <5 μg/ml to a top value of >65 μg/ml if the buffer volume is 2.5 ml (as in a UV cuvette). If it is clear that precipitation is occurring early in the addition sequence, we stop the addition so that we have two consecutive readings after the precipitate is first detected. Precipitation can be quantified by an absorbance increase due to light scattering by precipitated particulate material in a dedicated diode array UV machine. The sensitivity to light scattering is a function of the placement of the diode array detector relative to the cuvette and differs among instruments. We found that the array placement in a Hewlett Packard HP8452A diode array gives high sensitivity to light scattering. Increased UV absorbance from light scattering is measured in the 600–820 nm range because most drugs have UV absorbance well below this range.

In its simplest implementation, the precipitation point is calculated from a bilinear curve fit to the Absorbance (y axis) vs. μl of DMSO (x axis) plot. The coordinates of the intersect point of the two line segments are termed X crit and Y crit. X crit is the microlitres of DMSO added when precipitation occurs and Y crit is the UV Absorbance at the precipitation point. The concentration of drug in DMSO (10 μg/ml) is known. The volume of aqueous buffer (typically 2.5 ml in a cuvette) is known so the drug concentration expressed as μg of drug per ml buffer at the precipitation point is readily calculated. The volume percent aqueous DMSO at the precipitation point is also reported. Under our assay conditions it does not exceed 0.67% for a turbidimetric solubility of >65 μg/ml. The upper solubility limit is based on the premise that for most projects permeability is not a major problem and that solubility assays will most often be requested for poorly soluble compounds. In the absence of poor permeability, solubilities above 65 μg/ml suggest that if bio-availability is poor, solubility is not the problem.

2.18. Technical considerations and signal processing

In our experience, most UV active compounds made in our Medicinal Chemistry labs have UV peak maxima below 400 nm. Approximation to a Gaussian form for absorbance peaks allows an estimate for the UV absorbance at long wavelength from the peak maximum and peak width at half height. A soluble compound with maximum absorbance at 400 nm and extinction coefficient of 10 000 and peak width at half height of 100 nm at a concentration of 400 μg/ml (well above the maximum for our assay) has calculated absorbance of 0.000151 at 600 nm.

The sensitivity of UV absorbance measurements to light scattering is largely a function of how closely the diode array is positioned to the UV cuvette and varies among manufacturers. The HP89532 DOS software detects a curve due to light scattering by fitting the absorbance over a wavelength range to a power curve of the form. $\text{Abs} = \text{k} \times \text{nm}^{-\text{n}}$, where k is a constant, nm = wavelength.

Values for ‘n’ were examined in a total of 45 solubility experiments. The last scan in each solubility series was examined since precipitation is most likely at the highest drug concentration. In this 45 assay series precipitation was not observed in 10 assays (as assessed by values of $n>0$). Positive values of n ranged as high as 5.054 in the 35 assays in which precipitation occurred. Once precipitation occurred, all scans in an assay sequence could be fit with a power curve. The overall absorbance increase due to light scattering can be quite low. In most of the 45 assays, the total absorbance increase at 690 nm (due to precipitate formation) was in the OD range 0–0.01. Half the absorbance increases were in the range 0–0.001. Measurements within these very small ranges quantitate the precipitation point.

Problems in determining the precipitation point occur when a compound is intensely colored since colored compounds may be miscalled as insoluble. In collaboration with Professor Chris Brown at the University of Rhode Island, we implemented a fast

fourier transform (FFT) signal processing procedure to enhance assay sensitivity and to avoid false positive solubility values due to colored compounds [20]. The absorbance curve due to light scattering has an apparent peak width at half height which is much wider than the apparent peak width at half height for a typical UV absorption curve. An analysis procedure that is sensitive to the degree of curvature can be used to differentiate color from light scattering. The even wavelength spacing in our diode array UV means that the absorbance vs. wavelength matrix in each scan can be treated as if it were a time series (which it really is not). In a time series, the early terms in an FFT describe components of low curvature (low frequency). An FFT over a 256 nm range (566–820 nm) generates 128 absorbance values which in turn generates 128 FFT terms. FFT term 1 describes the baseline shift. By plotting the real component of FFT term 1 or term 2 vs. DMSO addition, the false positive rate from color is much reduced and we detect the onset of precipitation as if we were plotting absorbance at a single wavelength vs. absorbance.

An alternative to the use of a dedicated diode array UV is to use one of a number of relatively inexpensive commercially available nephelometers. The solubility protocol using a nephelometer as the signal detector is identical to that using a UV machine. We have experience using a HACH AN2100 as a turbidity detector. A nephelometer has the advantage that colored impurities do not cause a false positive precipitation signal and so signal processing is avoided. The disadvantage is the larger volume requirement relative to a UV cuvette. The HACH unit uses inexpensive disposable glass test tubes that can be as small as 100 mm $\times$ 12 mm. The use of even smaller tubes and the resultant advantage of reduced volume is precluded by light scattering from the more sharply curved surface of a smaller diameter tube.

Using nephelometric turbidity unit (NTU) standards, the threshold for detection using a UV detector-based assay is 0.2 NTUs and a 0.4 NTU standard can be reliably detected vs. a water blank. Turbidity standards in the range 0.2–2 NTU units suffice to cover the scattering range likely to be detected in a solubility assay. Some type of signal detector is necessary if light scattering is the analytical signal used to detect precipitation. For example, a 1.0 NTU standard was our lower visual detection limit using a fiber optic illuminator to visualize Tyndall light scattering. The European Pharmacopoeia defines the lowest category of turbidity — 'slight opalescence' on the basis of measured optical density changes in the range 0.0005–0.0156 at 340–360 nm. These optical density readings correspond to NTU standards well below 1.0 (in the 0.2–0.4 range) in our equipment.

3. Calculation of absorption parameters

3.1. Overall approach

The four parameters used for the prediction of potential absorption problems can be easily calculated with any computer and a programming language that supports or facilitates the analysis of molecular topology. At Pfizer, we began our programming efforts using MDL's sequence and MEDIT languages for MACCS and have since successfully ported the algorithms to Tripos' SPL and MDL's ISIS PL languages without difficulty.

The parameters of molecular weight and sum of nitrogen and oxygen atoms are very simple to calculate and require no further discussion. Likewise, the calculation of the number of hydrogen-bond acceptors is simply the number of nitrogen and oxygen atoms attached to at least one hydrogen atom in their neutral state.

3.2. MLogP. Log P by the method of Moriguchi

The calculation of log P via the method of Moriguchi et al. [11] required us to make some assumptions that were not clear from the rules and examples in the two papers describing the method [11,12]. Therefore, more detailed discussion on how we implemented this method is necessary.

The method begins with a straightforward counting of lipophilic atoms (all carbons and halogens with a multiplier rule for normalizing their contributions) and hydrophilic atoms (all nitrogen and oxygen atoms). Using a collection of 1230 compounds, Moriguchi et al. found that these two parameters alone account for 73% of the variance in the

experimental log Ps. When a 'saturation correction' is applied by raising the lipophilic parameter value to the 0.6 power and the hydrophilic parameter to the 0.9 power, the regression model accounted for 75% of the variance.

The Moriguchi method then applies 11 correction factors, four that increase the hydrophobicity and seven that increase the lipophilicity, and the final equation accounts for 91% of the variance in the experimental log Ps of the 1230 compounds. The correction factors that increase hydrophobicity are:

1. UB, the number of unsaturated bonds except for those in nitro groups. Aromatic compounds like benzene are analyzed as having alternating single and double bonds so a benzene ring has 3 double bonds for the UB correction factor, naphthalene has a value of 5;
2. AMP, the correction factor for amphoteric compounds where each occurrence of an alpha amino acid structure adds 1.0 to the AMP parameter, while each amino benzoic acid and each pyridine carboxylic acid occurrence adds 0.5;
3. RNG, a dummy variable which has the value of 1.0 if the compound has any rings other than benzene or benzene condensed with other aromatic, hetero-aromatic, or hydrocarbon rings;
4. QN, the number of quaternary nitrogen atoms (if the nitrogen is part of an N-oxide, only 0.5 is added).

The seven correction factors that increase lipophilicity are:

1. PRX, a proximity correction factor for nitrogen and oxygen atoms that are close to one another topologically. For each two atoms directly bonded to each other, add 2.0 and for each two atoms connected via a carbon, sulfur, or phosphorus atom, add 1.0 unless one of the two bonds connecting the two atoms is a double bond, in which case, according to some examples in the papers, you must add 2.0. In addition, for each carboxamide group, we add an extra 1.0 and for each sulfonamide group, we add 2.0;
2. HB, a dummy variable which is set to 1.0 if there are any structural features that will create an internal hydrogen bond. We limited our programs to search for just the examples given in the Moriguchi paper [11] as it is hard to determine how strong a hydrogen bond has to be to affect lipophilicity;
3. POL, the number of heteroatoms connected to an aromatic ring by just one bond or the number of carbon atoms attached to two or more heteroatoms which are also attached to an aromatic ring by just one bond;
4. ALK, a dummy parameter that is set to 1.0 if the molecule contains only carbon and hydrogen atoms and no more than one double bond;
5. NO2, the number of nitro groups in the molecule;
6. NCS, a variable that adds 1.0 for each isothiocyanate group and 0.5 for each thiocyanate group;
7. BLM, a dummy parameter whose value is 1.0 if there is a beta lactam ring in the molecule.

3.3. MLogP calculations

Log Ps, calculated by our Moriguchi-based computer program for a set of 235 compounds were less accurate than the calculated log Ps (CLogPs) from Hansch and Leo's Pomona College Medicinal Chemistry Project MedChem software distributed by Biobyte. The set of 235 was chosen so that the CLogP calculation would not fail because of missing fragments. Our implementation of the Moriguchi method accounts for 83% of the variance with a standard error of 0.6 whereas the Hansch values account for 96% of the variance with a standard error of 0.3. The advantages of the Moriguchi method are that it can be easily programmed in any language so that it can be integrated with other systems and it does not require a large database of parameter values.

4. The development setting: prediction of aqueous thermodynamic solubility

4.1. General considerations

The prediction of the aqueous solubility of drug candidates may not be a primary concern in early screening stages, but the knowledge of the thermodynamic solubility of drug candidates is of paramount importance in assisting the discovery, as

well as the development, of new drug entities at later stages. A poor aqueous solubility is likely to result in absorption problems, since the flux of drug across the intestinal membrane is proportional to its concentration gradient between the intestinal lumen and the blood. Therefore even in the presence of a good permeation rate a low absorption is likely to be the result. Conversely, a compound with high aqueous solubility might be well absorbed, even if it possesses a moderate or low permeation rate.

Formulation efforts can help in addressing these problems, but there are severe limitations to the absorption enhancement that can be realistically achieved. Stability and manufacturing problems also have to be taken into account since it is likely that an insoluble drug candidate may not be formulated as a conventional tablet or capsule, and will require a less conventional approach such as, for example, a soft gel capsule. Low solubility may have an even greater impact if an i.v. dosage form is desired. Obviously, a method for predicting solubility of drug candidates at an early stage of discovery would have a great impact on the overall discovery and development process.

Unfortunately the aqueous solubility of a given molecule is the result of a complex interplay of several factors ranging from the hydrogen-bond donor and acceptor properties of the molecule and of water, to the energetic cost of disrupting the crystal lattice of the solid in order to bring it into solution (‘fluidization’) [24].

In any given situation, not all the factors may play an important role and it is difficult to predict the solubility of a complex drug candidate, on the basis of the presence or absence of certain functional groups. Conformational effects in solution may play a major role in the outcome of the solubility and cannot be accounted for by a simple summation of ‘contributing’ groups.

Thus, any method which would aim at predicting the aqueous solubility of a given molecule would have to take into account a more comprehensive ‘description’ of the molecule as the outcome of the complex interplay of factors.

The brief discussion of the problem outlined above can be summarized by considering the three basic quantities governing the solubility (S) of a given solid solute:

$$S = f(\text{Crystal Packing Energy} + \text{CavitationEnergy} + \text{Solvation Energy})$$

In this equation, the crystal packing energy is a (endoergic) term which accounts for energy necessary to disrupt the crystal packing and to bring isolated molecules in gas phases, i.e. its enthalpy of sublimation. The cavitation energy is a (endoergic) term which accounts for the energy necessary to disrupt water (structured by its hydrogen bonds) and to create a cavity into which to host the solute molecule. Finally, the solvation energy might be defined as the sum (exoergic term) of favorable interactions between the solvent and the solute.

In dealing with the prediction of the solubility of crystalline solids[2], a first major hurdle to overcome is the determination or estimation of their melting point or, better, of their enthalpy of sublimation. At present no accurate and efficient method is available to predict these two quantities for the relatively complex molecules which are encountered in the pharmaceutical research. Gavezzotti[3] [26] has discussed this point in a review article on the predictability of crystal structures and he states that ‘...the melting point is one of the most difficult crystal properties to predict.’ This author has pioneered the use of computational methods to predict crystal structures and polymorphs and, consequently, properties such as melting point and enthalpy of sublimation. A commercially available program has been recently developed [27] but the use of these approaches is still far from being routine and from being useful in a screening stage for a relatively large number of compounds, all of which possess a relatively high conformational flexibility.

Thus, although there are several approaches to estimating and predicting the solubility of organic compounds, the authors of this article are of the opinion that none of the presently available methods can truly be exploited for a relatively accurate

[2]Since the vast majority of drug molecules and most substances of pharmaceutical interest are crystalline solids, this discussion will focus on the prediction of the solubility of crystalline solids.

[3]The program PROMET is available from Professor Gavezzotti, University of Milan, Italy.

prediction of solubility, if the target of the prediction is the solubility of complex pharmaceutical drug candidates. Although the judicious application of some these approaches might be useful for 'rank-ordering' of compounds and prioritization of their synthesis, we are not aware of any such systematic use of estimation methods.

The sections that follow will discuss available methods, taking into account the second and third terms of the above relationship and the feasibility of their assessment a priori, and they will be treated as one term since the available methods consider the interactions in solution as the (algebraic) sum of the two terms and their contributors. This discussion is by no means exhaustive but it is rather intended as an overview of the methods available as seen, in particular, from a pharmaceutical perspective.

4.2. LSERs and TLSER methods

Linear Solvation Energy Relationships (LSERs), based upon solvatochromic parameters, have the advantage of a good theoretical background and offer a correlation between several molecular properties, and a solute property, SP. Several LSERs have been developed over the past few years and they seem to work well for predicting a generalized SP for a series of solutes in one or more (immiscible) phases. Most notably, the work of Abraham et al. [28] has generated an equation of the general type:

$$LogSP = c + rR_2 + a\Sigma\alpha_2^H + b\Sigma\beta_2^H + s\pi_2^H + nV_x$$

where c is a constant, R_2 is an excess molar refractivity, $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ are the (summation or 'effective') solute hydrogen-bond acidity and basicity, respectively, π_2^H is the solute dipolarity-polarizability and V_x is McGowan's characteristic volume [29]. The main problem encountered when using parameterized equations is that such quantities (parameters or descriptors) cannot easily be estimated, from structures only, for complex multi-functional molecules such as drug candidates, especially if they are capable of intra-molecular hydrogen bonding, as is often the case. Nevertheless, the method was successfully applied to the correlation between the solvatochromic parameters described above and the aqueous solubility of relatively simple organic non-electrolytes [30].

More recently, Kamlet [31] has published equations describing the solubility of aromatic solutes including polycyclic and chlorinated aromatic hydrocarbons. In these equations a term accounting for the crystal packing energy was introduced, and the equation has the general form:

$$\log S_w(aromatics) = \frac{0.24 - 5.28V_I}{100} + 4.03\beta_m + 1.53\alpha_m - 0.0099(m.p. - 25)$$

where V_I is the intrinsic (van der Waals) molar volume of the solute, the other parameters are defined as above and the subscript *m* indicates a non self-associating solute monomer. It is interesting to note that the term 0.0099(m.p.−25) is used, in the words of the author, 'to account for the process of conversion of the solid solute to super-cooled liquid at 25°C.' This term is therefore related to the crystal packing energy mentioned earlier, albeit representing the conversion from a solid to a 'super-cooled' liquid, not to isolated molecules in gas phase. The author finds the above term 'robust' in its statistical significance and it should be noted that coefficient of 0.0099 implies that a variation of less than one order of magnitude will be observed for variations in melting points of less than 100°C.

This finding might be exploited in a series of close structural analogs where a large variation in melting points (>100°C) is not expected (as might often be the case) and the 'solution behavior' could be estimated by solvatochromic parameters. Thus, with some error, the prioritization of more soluble synthetic targets might be achieved, since the relative ('rank-order') solubility of structurally close analogs may be all that it is sought at an early stage. However this prioritization would rely on the assumption that variations in structural properties which bring about a (desired) lowering of the crystal packing energy, would not significantly and adversely alter the properties of a molecule with respect to its solvation in water. If the lower crystal packing energy is the result, for example, of a lower hydrogen-bond capability, a diminished solvation in water may offset the lowering of the crystal packing energy.

Even with the assumption described above, the estimation of a relatively good rank-ordering of aqueous solubilities would still require the determination of solvatochromic parameters which is generally achieved through the determination of several partition coefficients. On the other hand, descriptor values for several fragments (functional groups) are available and they may be used to calculate the 'summation' parameters for the molecules of interest. This process is not without caveats though, as a very judicious choice of the 'disconnection pattern' must be made to obtain reliable results. In a recent paper describing the partition of solutes across the blood-brain barrier, Abraham et al. [32] reported the calculation and use of these descriptors for compounds of pharmaceutical interest but he warned about the possibility of inter-molecular hydrogen bonding, which may be a source of error if not present in the 'reference' compounds, and pointed out the fact that these correlations are best used within the descriptors range used to generate them.

Some authors have reported the calculation of quantities related to those descriptors, via ab initio [33–35] or semi-empirical methods [36,37]. The equations stemming from computed values have been termed TLSERs (Theoretical Linear Solvation Energy Relationships) [36]. However, we are not aware of any application of this approach to a series of complex multifunctional compounds, and these types of correlations are likely to be difficult for these compounds, due to the relatively high level of computation involved.

Ruelle and Kesselring and colleagues [38–40] reported a multi-parameter equation, qualitatively similar to the LSERs described above. This equation attempts to predict solubility by using terms which account for the quantities that play a role in the process. It does contain a solute 'fluidization' term (endoergic cost of destroying the crystal lattice of a solid) and other terms describing the hydrophobic effect, hydrogen bond formation between proton-acceptor solutes and proton-donor solvents, and the H-bond formation between amphiphilic solutes and proton acceptor and/or proton-donor solvents as well as the auto-association of the solute in solution.

Although this equation takes into account the free energy changes involved in the dissolution process, in our opinion its complexity prevents its use for multifunctional molecules. The examples reported address simple hydrocarbons or mono-functional molecules and much emphasis is placed on organic (associated and non-associated) solvents. In many such cases, approximations leading to the cancellation of some term, can be made but, if an attempt to predict the solubility of complex drug candidates in water is made, all those terms might be present at the same time and thus it would be very difficult to treat solubility within the framework of this equation.

4.3. LogP and AQUAFAC methods

Prominent in this area is the work of Yalkowski [41] who has published a series of papers describing the prediction of solubility using LogP (the logarithm of the octanol/water partition coefficient) and a term describing the energetic cost of the crystal lattice disruption. However Yalkowski's work is largely based on the prediction or estimation of the solubility of halogenated aromatic and polycyclic halogenated aromatic hydrocarbons [42], due to their great environmental importance. The general solubility equation, for organic non-electrolytes is reported below.

$$\log S_{pred} = -\frac{\Delta S_m(m.p. - 25)}{1364} - \log P + 0.80$$

In this equation, ΔS_m is the entropy of melting and m.p. is the melting point in °C. The signs of the two terms considered are physically reasonable, since an increase in either the first term (higher crystal packing energy) or in LogP (more lipophilic compound), would cause a decrease in the observed (molar) solubility S_m. In a recent paper [43], this author discusses the predictive use of the above equation and, in particular, the prediction of activity coefficients. The latter is a term which accounts for deviations from ideal solubility behavior due to differences in size and shape, but also in hydrogen bonding ability, between the solute and the solvent. The conclusion is that, among methods based upon solvatochromic parameters, or simply based on molecular volume, molecular weight or regular solution theory, the estimation of the activity coefficient is best achieved by using the LogP method.

Many computational methods are indeed available to address the prediction of LogP and the aqueous solubility of complex molecules. A well known and widely used program to predict LogP values is

CLogP [44] which uses a group-contribution approach to yield a LogP value. Another method, developed by Moriguchi et al. [11], which uses atomic constants and correction factors to account for different atom types is discussed in detail in Section 3.2. We have observed that, in the daily practice of pharmaceutical sciences, both methods have their 'outliers' but methods based on fragmental constants tend to fail, in the not infrequent instances where appropriate constants are not available.

However, LogP prediction aside, the method reported by Yalkowski was developed on a data set largely based upon rigid, polycyclic and halogenated aromatic compounds and does not seem to easily yield itself to the prediction of complex pharmaceutical compounds. The basic difficulty is that while LogP could be estimated albeit with some error by computational approaches, the melting point and entropy of melting are still difficult to calculate or even simply to estimate. Yalkowski discusses this point in several papers [42,45,46] and shows the relationship between the entropy of fusion and the molecular rotational and translational entropies. Some rules are offered for the estimation of entropy, but the work is limited to relatively simple molecules. The melting point prediction is also discussed and a computational approach, based on molecular properties such as eccentricity (the ratio between the maximum molecular length and the mean molecular diameter) is proposed. However, the calculation of such properties may be easy to perform on simple polychlorinated biphenyls, but would not easily be applicable for complex drug candidates.

A similar approach to solubility predictions using a group-contribution method has been implemented in the CHEMICALC-2 program [47], which calculates LogP and log $1/S$ where S is the molar aqueous solubility. This program uses several different algorithms to calculate log $1/S$ depending on the complexity and nature of the molecule, and requires knowledge of the melting point, T_m. If T_m is not available, the program calculates the solubility of the super-cooled liquid at 25°C. In the case of complex molecules, fragmental constants may be missing from its database and poor results are obtained. We have used this program to some extent and we are not encouraged by the correlation between 'predicted' and experimental solubility.

Yalkowski and colleagues [48] have more recently discussed an improvement of the AQUAFAC (AQUeous Functional group Activity Coefficients) fragmental constant method. In this work, the authors describe a correlation between the sum of fragmental constants of a given molecule and the activity coefficient, defined as a measure of the non-ideality of the solution. The knowledge or estimation of ΔS_m and m.p. is necessary, but the method seems to be somewhat better than the general solubility equation based on LogP values. Yalkowski explains this by pointing out that these group contribution constants were derived entirely from aqueous phase data and they should perform better than octanol-water partition coefficients. We concur with this explanation since it is known that the octanol-water partition coefficients are rather insensitive to the hydrogen-bond donor capability of the solute. Furthermore, the authors point out the fact that molecules like small carboxylic acids are likely to dimerize in octanol, while in water they would not.

The solubility equation derived using the AQUAFAC coefficients is reported below.

$$\log S_{pred} = -\frac{\Delta S_m(m.p. - 25)}{1364} - \Sigma n_i q_i$$

where q_i is the group contribution of the ith group and n_i is the number of times the ith group appears in the molecule. The negative sign of the second term stems from the fact that the constant of polar groups (e.g. OH = −1.81) has a negative sign and a net negative sign of the summation of contributors would yield an overall positive contribution to solubility. However, while this method might be of simple application, its scope seems limited to molecule containing relatively simple functional groups, and the objections to the use of group contribution methods, which do not consider conformational effects, remain.

4.4. Other calculation methods

Bodor and Huang [49] and Nelson and Jurs [50] have reported methods based entirely on calculated geometric, electronic and topological descriptors, for a series of relatively simple liquid and solid solutes.

We favor these methods as truly a priori predic-

tions based on molecular structures only, but some questions arise when the compounds have conformational flexibility and multiple functional groups, and some of the descriptors will depend upon the particular conformation chosen. As it is generally true for many QSAR approaches, there is uncertainty about the actual predictive value of a test set which does not include a wide variety of compounds and, in Bodor's training set of 331 compounds we fail to recognize with few exceptions represented by rigid steroids, complex multifunctional molecules. Furthermore a large number of the compounds used are liquids or gases at ambient temperature.

Bodor's method involves the calculation of 18 descriptors, among which are the ovality of the molecule, the calculated dipole moment, and the square root of the sum of squared charges on oxygen atoms, but it does yield a good correlation for the 331-compound set. The predictive power of the model is illustrated by a table of 17 compounds, but most of them are rigid aromatics, although a reasonably good prediction is offered for dexamethasone. The latter however is an epimer of betamethasone which is present in the training set, and it is difficult to predict the robustness of the correlation with regard to its application to a truly diverse set of molecules. Similar considerations could be extended to the work by Nelson and Jurs, which is also based on calculated descriptors and it does not seem to involve any polyfunctional molecule or any solid compound at 25°C. Overall the correlation is good but the caveats on its application to drug-like compounds remain, as well as our objections on the ease of calculation of the parameters for compounds of pharmaceutical interest.

Finally, Bodor et al. [25] and Yalkowski and colleagues [5] have reported the use of neural networks to develop correlations using the calculated parameters discussed above or the AQUAFAC coefficients, respectively. While we have no direct experience with the use of neural networks, we are of the opinion that it may not be a trivial task to set up and 'train' a neural network and the superiority of this approach in comparison to 'conventional' regression techniques may be more apparent than real. Indeed Bodor reports a similar standard deviation for the prediction using the neural network or regression analysis [49] on the same data set, and the use of a neural network does not appear to offer any advantage over the regression analysis.

5. Conclusion

Combinatorial chemistry and high throughput screening (HTS) techniques are used in drug research because they produce leads with an efficiency that compares favorably with 'rational' drug design and, perhaps more importantly, because these techniques expand the breadth of therapeutic opportunities and hence the leads for drug discovery. Established methodology allows the medicinal chemist, often in a relatively short time, to convert these novel leads to compounds with in vitro potency suitable to a potential drug candidate. This stage of the discovery process is highly predictable. However, the majority of drugs are intended for oral therapy and introducing oral activity is not predictable, is time and manning expensive and can easily consume more resources than the optimization of in vitro activity. The in vitro nature of HTS screening techniques on compound sets with no bias towards properties favorable for oral activity coupled with known medicinal chemistry principles tends to shift HTS leads towards more lipophilic and therefore generally less soluble profiles. This is the tradeoff in HTS screening. Efficiency of lead generation is high, and therapeutic opportunities are much expanded, but the physical profiles of the leads are worse and oral activity is more difficult. Obtaining oral activity can easily become a rate-limiting step and hence methods which allow physico-chemical predictions from molecular structure are badly needed in both early discovery and pharmaceutical development settings.

Computational methods in the early discovery setting need to deal with large numbers of compounds and serve as filters which direct chemistry SAR towards compounds with greater probability of oral activity. These computational methods become particularly important as experimental studies become more difficult because compounds are available for physico-chemical screening in only very small quantities and in non-traditional formats. Early discovery methods deal with probabilities and not exact value predictions. They enhance productivity

by indicating which types of compounds are less likely to be absorbed and which are more likely to require above average manning expenditures to become orally active. Calculations, however imprecise, are better than none when choices must be made in the design or purchase of combinatorial libraries. Drug discovery requires a starting point — a lead. Hence the current literature correctly focuses on improving in vitro activity detection by optimizing chemical diversity so as to maximize coverage of three-dimensional receptor space. Assuming this goal is not compromised by physico-chemical calculations, we believe a competitive advantage accrues to the organization that can identify compound sets likely to give leads more easily converted to orally active drugs.

Methods in the pharmaceutical developmental setting deal with much smaller numbers of compounds. Here, a more accurate prediction is computationally complex because exact values rather than probabilities are important, and because the prediction of crystal packing energies is at present extremely difficult. The problem of polymorphism, common in pharmaceutical research, which may have been deferred in the discovery setting has to be addressed in the development setting. Currently, only approximate estimates of the solubility of multifunctional and conformationally flexible drug candidates are possible and these need to be supported by physical measurements which provide experimental 'feedback' on analogs in a particular class of compounds. In our view, a priori solubility estimation methods like Bodor's multi-parameter equation [49] are the current best choice, but some of the required properties are not easily computed without a preliminary optimization of preferred conformations and good initial estimates. The accurate prediction of the solubility of complex multifunctional compounds at the moment still remains an elusive target. The requirements for high accuracy and the complexity of possible studies in the drug developmental setting means that even small changes towards poorer, but still acceptable, physico-chemical properties in compounds approaching candidacy can translate to higher developmental time and manning requirements. Moreover, there has not been the same level of efficiency improvement in many developmental assays as there has been in discovery screening. For example, there is not the same level of efficiency improvement in measuring accurate equilibrium solubility as there has been in the efficiency of detecting leads.

Medicinal chemists efficiently and predictably optimize in vitro activity, especially when the lead has no key fragments missing. This ability will likely be reinforced because the current focus on chemical diversity should produce fewer leads with missing fragments. Oral activity prospects are improved through increased potency, but improvements in solubility or permeability can also achieve the same goal. Despite increasingly sophisticated formulation approaches, deficiencies in physico-chemical properties may represent the difference between failure and the development of a successful oral drug product.

References

[1] Kubinyi, H. (1995) Strategies and recent technologies in drug discovery. Pharmazie 50, 647–662.
[2] Moos, W.H. and Green, G.D. (1993) Chapter 33. Recent Advances in the Generation of Molecular Diversity. Annu. Rep. Med. Chem. 28, 315–324.
[3] Patel, D.V. and Gordon, E.M. (1996) Applications of small-molecule combinatorial chemistry to drug discovery. Drug Discov. Today 1, 134–144.
[4] Baum, R.M. (1994) Combinatorial approaches provide fresh leads for medicinal chemistry. Chem. Eng. News February 7, 20–26.
[5] Chow, H., Chen, H., Ng, T., Nyrdal, P. and Yalkowski, S.H. (1995) Using backpropagation networks for the estimation of aqueous activity coefficients of aromatic compounds. J. Chem. Inf. Comput. Sci., 35, 723–728.
[6] Andrews, P.R., Craik, D.J. and Martin, J.L. (1984) Functional group contributions to drug-receptor interactions. J. Med. Chem. 27, 1648–1657.
[7] Navia, M.A. and Chaturvedi, P.R. (1996) Design principles for orally bioavailable drugs. Drug. Dev. Today 1, 179–189.
[8] Pardridge, W.M. (1995) Transport of small molecules through the blood-brain barrier: biology and methodology. Adv. Drug Deliv. Rev. 15, 5–36.
[9] Cohen, B.E. and Bangham, A.D. (1972) Diffusion of small non-electrolytes across liposome membranes. Nature 236, 173–174.
[10] Testa, B., Carrupt, P-A., Gaillard, P., Billois, F. and Weber, P. (1996) Lipophilicity in molecular modelling. Pharm. Res. 13, 335–343.
[11] Moriguchi, I., Hirono, S., Liu, Q., Nakagome, Y. and Matsushita, Y. (1992) Simple method of calculating octanol/water partition coefficient. Chem. Pharm. Bull. 40, 127–130.

[12] Moriguchi, I., Hirono, S., Nakagome, I. and Hirano, H. (1994) Comparison of log P values for drugs calculated by several methods. Chem. Pharm. Bull. 42, 976–978.
[13] Leo, A.J. (1995) Critique of recent comparison of log P calculation methods. Chem. Pharm. Bull. 43, 512–513.
[14] Abraham, M.H., Chadha, S.H., Whiting, G.S. and Mitchell, R.C. (1994) Hydrogen bonding. 32. An analysis of water-octanol and water-alkane partitioning and the delta log P parameter of Seiler. J. Pharm. Sci. 83, 1085–1100.
[15] Paterson, D.A., Conradi, R.A., Hilgers, A.R., Vidmar, T.J. and Burton, P.S. (1994) A non-aqueous partitioning system for predicting the oral absorption potential of peptides. Quant. Struct.-Act. Relatsh. 13, 4–10.
[16] Hillgren, K.M., Kato, A. and Borchardt, R.T. (1995) In vitro systems for studying intestinal drug absorption. Med. Res. Rev. 15, 83–109.
[17] Abraham, M.H. (1993) Hydrogen bonding. 31. Construction of a scale of solute effective or summation hydrogen-bond basicity. J. Phys. Org. Chem. 6, 660–684.
[18] Raevsky, O.A., Grigor'ev, V.Y., Kireev, D.B. and Zefirov, N.S. (1992) Complete thermodynamic description of H-bonding in the framework of multiplicative approach. Quant. Struct.-Act. Relatsh. 11, 49–63.
[19] Raevsky, O.A., Schaper, K-J. and Seydel, J.K. (1995) H-Bond contribution to octanol-water partition coefficients of polar compounds. Quant. Struct.-Act. Relatsh. 14, 433–436.
[20] Lipinski, C.A. (1995) Computational alerts for potential absorption problems: profiles of clinically tested drugs. In: Tools for Oral Absorption. Part Two. Predicting Human Absorption. BIOTEC, PDD symposium, AAPS, Miami.
[21] Prediction of Drug Transport Through Membranes. Chicago IAM Meeting, August 18, 1995.
[22] Workshop on Oral Drug Delivery: Interface Between Discovery and Development. Report and Recommendations. National Institutes of Health, National Institute of General Medical Sciences. Herndon, Virginia, December 5–7, 1993.
[23] Lipinski, C.A., Davis, A.L., Aldridge, P.K. and Brown, C.W. (1994) Biologically relevant UV-based solubility assays. Pharmaceutical Analysis Session, 45th Pittsburgh Conf., Chicago, Oral Presentation 959.
[24] Reichardt, C. (1988) Solvents and Solvent Effects in Organic Chemistry. 2nd edn. VCH, Weinheim, pp. 27–35.
[25] Bodor, N., Harget, A. and Huang, N.-J. (1991) Neural network studies. 1. Estimation of the aqueous solubility of organic compounds. J. Am. Chem. Soc. 113, 9480–9483.
[26] Gavezzotti, A. (1994) Are crystal structures predictable? Acc. Chem. Res. 27, 309–314.
[27] Polymorph Predictor, Molecular Simulations Inc., Burlington, MA.
[28] Abraham, M.H. (1993) Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes. Chem. Soc. Rev. 22, 73–83.
[29] Abraham, M.H. and McGowan, J.C. (1987) The use of characteristic volumes to measure cavity terms in reversed phase liquid chromatography. Chromatographia 23, 243–246.
[30] Taft, R.W., Abraham, N.H., Doherty, R.M. and Kamlet, M.J. (1985) The molecular properties governing solubilities of organic nonelectrolytes in water. Nature 315, 384–386.
[31] Kamlet, M.J. (1993) Linear solvation energy relationships: an improved equation for correlation and prediction of aqueous solubilities of aromatic solutes including polycyclic aromatic hydrocarbons and polychlorinated biphenyls. Prog. Phys. Org. Chem. 19, 295–317.
[32] Abraham, M.H., Chadha, H.S. and Mitchell, R.C. (1994) Hydrogen bonding. 33. Factors that influence the distribution of solutes between blood and brain. J. Pharm. Sci. 83, 1257–1268.
[33] Brinck, T., Murray, J.S. and Politzer, P. (1993) Octanol/water partition coefficients expressed in terms of solute molecular surface areas and electrostatic potentials. J. Org. Chem. 58, 7070–7073.
[34] Murray, J.S. and Politzer, P. (1991) Correlations between the solvent hydrogen-bond-donating parameter α and the calculated molecular surface electrostatic potential. J. Org. Chem. 56, 6715–6717.
[35] Murray, J.S., Ranganathan, S. and Politzer, P. (1991) Correlations between the solvent hydrogen bond acceptor parameter β and the calculated molecular electrostatic potential. J. Org. Chem. 56, 3734–3737.
[36] Famini, G.R., Penski, C.A. and Wilson, L.Y. (1992) Using theoretical descriptors in quantitative structure activity relationships: some physicochemical properties. J. Phys. Org. Chem. 5, 395–408.
[37] Headley, A.D., Starnes, S.D., Wilson, L.Y. and Famini, G.R. (1994) Analysis of solute/solvent interactions for the acidity of acetic acids by theoretical descriptors. J. Org. Chem. 59, 8040–8046.
[38] Ruelle, P., Sarraf, E. and Kesselring, U.W. (1994) Prediction of carbazole solubility and its dependence upon the solvent nature. Int. J. Pharm. 104, 125–133.
[39] Ruelle, P., Rey-Mermet, C., Buchmann, M., Nam-Tran, H., Kesselring, U.W. and Huyskens, P.L. (1991) A new predictive equation for the solubility of drugs based on the thermodynamics of mobile disorder. Pharm. Res. 8, 840–850.
[40] Ruelle, P. and Kesselring, U.W. (1994) Solubility predictions for solid nitriles and tertiary amides based on the mobile order theory. Pharm. Res. 11, 201–205.
[41] Yalkowski, S.H. and Valvani, S.C. (1980) Solubility and partitioning 1: solubility of nonelectrolytes in water. J. Pharm. Sci. 69, 912–922.
[42] Abramovitz, R. and Yalkowski, S.H. (1990) Estimation of aqueous solubility and melting point of PCB congeners. Chemosphere 21, 1221–1229.
[43] Yalkowski, S.H. and Pinal, R. (1993) Estimation of the solubility of complex organic compounds. Chemosphere 26, 1239–1261.
[44] Leo, A. (1993) Calculating log Poct from structures. Chem. Rev. 93, 1281–1306 (CLOGP version 3.5).
[45] Abramovitz, R. and Yalkowski, S.H. (1990) Melting point, boiling point, and symmetry. Pharm. Res. 7, 942–947.
[46] Dannenfelser, R.M., Surendran, N. and Yalkowski, S.H. (1993) Molecular symmetry and related properties. SAR & QSAR Environ. Res. 1, 273–292.

[47] Suzuki, T. (1991) Development of an automatic estimation system for both the partition coefficient and aqueous solubility. J. Comput. Aided Mol. Des. 5, 149–166 (CHEMICALC-2, v. 1.0).
[48] Myrdal, P.B., Manka, A.M. and Yalkowsky, S.H. (1995) Aquafac 3: aqueous functional group activity coefficients; application to the estimation of aqueous solubility. Chemosphere 30, 1619–1637.
[49] Bodor, N. and Huang, M.-J. (1992) A new method for the estimation of the aqueous solubility of organic compounds. J. Pharm. Sci. 81, 954–959.
[50] Nelson, T.M. and Jurs, P.C. (1994) Prediction of aqueous solubility of organic compounds. J. Chem. Inf. Comput. Sci. 34, 601–609.