advanced
drug delivery
reviews

iews 46 (2001) 3–26

www.elsevier.com/locate/drugdeliv

ional approaches to estimate rug discovery and development ngs☆

ardo, Beryl W. Dominy, Paul J. Feeney

er Inc., Groton, CT 06340, USA

accepted 14 August 1996

ubility and permeability in discovery and development settings that poor absorption or permeation is more likely when there olecular weight (MWT) is greater than 500 and the calculated putational methodology for the rule-based Moriguchi Log P measurement is described and applied to known drugs. High and Log P and lower turbidimetric solubility than leads in the ns focus on exact value prediction and are difficult because of ships and Log P approaches are critically reviewed. Useful when coupled with experimental thermodynamic solubility erved.

eation; MWT; MLogP; H-Bond donors and acceptors; Turbidimetric

originally published in *Advanced Drug Delivery Reviews* 23 (1997)

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.

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.

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

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

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

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

(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

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

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

[c]	MWT	N+O[d]	Alert[e]
	225.21	8	0
	308.77	4	0
	180.16	4	0
	266.34	5	0
	749.00	14	1
	267.25	9	0
	334.40	6	0
	194.19	6	0
	515.65	8	0
	217.29	4	0
	236.28	3	0
	323.14	7	0
	252.34	6	0
	230.10	3	0
	1202.64	23	1
	266.39	2	0
	392.47	5	0
	284.75	3	0
	296.15	3	0
	414.53	6	0
	543.53	12	1
	376.46	7	0
	733.95	14	1
	337.45	9	0
	384.26	5	0
	130.08	4	0
	244.27	2	0
	330.75	7	0
	75.07	3	0
	297.74	7	0
	206.29	2	0
	280.42	2	0
	705.65	12	1
	380.92	1	0
	254.29	3	0
	328.42	5	0
	405.50	8	0
	182.18	6	0
	454.45	13	1
	267.37	4	0
	309.41	5	0
	327.38	5	0
	230.27	3	0
	263.39	1	0
	267.25	9	0
	451.49	10	0
	331.35	7	0
	383.41	9	0
	259.35	3	0
	324.43	4	0
	314.41	7	0
	303.36	5	0
	320.76	5	0
	471.69	3	0
	288.43	2	0
	416.36	7	0
	144.22	2	0
	811.00	13	1
	412.95	5	0

cted; 1, poor absorption or permeation are more likely.

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.

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-

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

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

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

lytical 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

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

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

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.

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.

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

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-

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

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.

(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.

[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.