Drugs of the Future 2002, 27(6): 577-588

CCC: 0377-8282/2002

Review Article

Computer-aided drug design: the role of quantitative structure-property, structure-activity and structure-metabolism relationships (QSPR, QSAR, QSMR)

Peter Buchwald[1] and Nicholas Bodor[1,2]*

*[1]IVAX Research, Inc., 4400 Biscayne Boulevard, Miami, FL 33137 and [2]Center for Drug Discovery, University of Florida, Gainesville, FL, USA. *Correspondence*

CONTENTS

Introduction

Drug design: quo vadis?

The discovery and development of a new chemical entity (NCE) that can reach the market as an effective new drug is a long, arduous and expensive process. Successful drug design still relies on the 4 Gs (Germans) noticed by Paul Ehrlich (1854-1915): "*Glück, Geduld, Geschick und Geld*" (that is luck, patience, skill and money). The odds of finding a new compound with the right combination of activity, selectivity, stability and safety are very unfavorable. Clinical studies now required to demonstrate statistically significant safety and efficacy are heavily regulated by government agencies and are expensive (*e.g.*, even phase I and II trials carry an average cost of in US$ 1-1.5 and 2-10 million, respectively). According to current estimates, from 30000 compounds synthesized, 2000 enter preclinical development, 200 enter phase I clinical trial, 40 enter phase II clinical trials, 12 enter phase III clinical trials, 8 are approved and only 1 makes a satisfactory return on investment (1). Despite strongly rising drug-related spending, which is mainly caused by the aging of the population and the rising quality of life in industrialized nations, the pharmaceutical industry is faced with increasing difficulties (2). Clinical trials alone can consume up to one half of the 20 year period reserved for exclusive rights to development and promotion, the main source of profit earning (3). Changes in legislature and medical insurance policy made it much easier for generic competitors, which operate on razor-thin margins, to cut into the market at the expiration of patent protection. New technologies and increasing regulatory demands have pushed research and development (R&D) expenditures close to 20% of sales from its value of 4% in 1966. Compare this to the current value of only 5-6% in the electronics industry, which is also heavily R&D-dependent (4).

Nevertheless, the pharmaceutical industry, as well as drug- and medicinal discovery as such, are in clear need of innovation (5, 6). Despite exponentially increasing R&D expenditures, the number of launched NCEs essentially stagnates (Fig. 1). Many highly trumpeted development failed to materialize in an increased NCE output. Furthermore, even among launched NCEs, there is considerable redundancy and very little real progress. For example, out of the 269 NCEs that received Food and Drug Administration (FDA) approval between 1976 and 1990, only an estimated 15% represented considerable improvement compared to 49% that represented no or negligible improvement and 35% that represented modest improvement (7).

In silico

According to our current understanding, the physiological effects generated by biologically active substances,

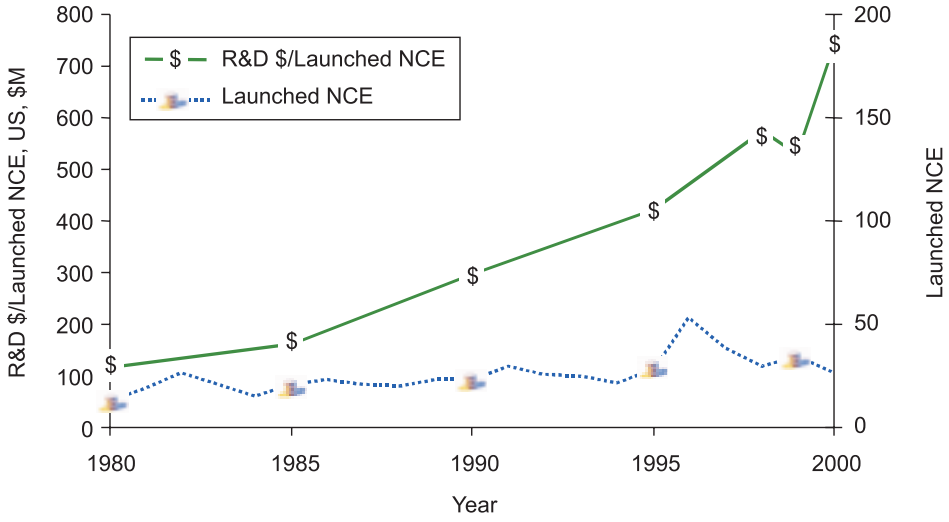

Fig. 1. The number of new chemical entities (NCEs, right vertical axis) launched annually in the United States was essentially stagnating whereas research and development (R&D) costs (shown here as calculated per launched NCE in million US$, left vertical axis) grew exponentially. (Graphic prepared based on data from ref. 124, 125.)

including drugs, are a function of the amount of active compound that actually reaches the "receptor", which in a general picture can be an enzyme, an ion channel, a receptor protein, a nucleic acid or any other biological macromolecule and a function of the strength of the interaction at this site (affinity) or the relevance of the structural changes produced. Because all these are ultimately determined by intermolecular forces, the main determinants of a compound's biological activity are its physicochemical properties and its relevant structural features. These also represent the basic assumptions underlying any (quantitative) structure-property, structure-activity, structure-metabolism or similar type of relationship (QSPR, QSAR, QSMR). Unfortunately, the observable final biological response is usually a function of many interacting, overlapping or competing processes from absorption, distribution, metabolism, elimination (ADME) to specific/nonspecific binding, induced structural changes and the presence of fine-tuned compensating biological mechanisms. Hence, a clear overall picture rarely emerges. Nevertheless, general guidelines for properties influencing passive distribution and related processes as well as structural requirements for certain well-defined molecular targets can often be derived. Because elimination of any unlikely candidate from chemical synthesis and from *in vitro*/*in vivo* testing provides significant time and financial savings, the development of reliable computerized (*in silico*) QSPR/QSAR/QSMR models or screening filters can afford considerable benefits. This is in fact the popular "fail early, fail cheap" strategy at its extreme. Because these fields widened considerably during the last years, space limitations allow us only a "review of reviews".

Computer-aided drug design

Structure-based design

Three-dimensional structures are now known for hundreds of proteins, which are or can become the target of drugs. Computer models comparing the structure of such a target with that of potential drug molecules can describe the probable interactions and save weeks, maybe months, of time plowing through laboratory tests. Solving the structure of bound ligand-receptor complexes, searching for potential leads that possess the required structural and chemical properties and then testing for activity (structure-based drug design) (8-10) has certainly been one of the quiet revolutions of the past years.

Property-based design

On the other hand, making a compound that binds with high affinity to a selected target into a successful and marketable drug can be very complicated and often impossible. Around 40% of putative drug candidates (NCEs) fail in their development because of inappropriate pharmacokinetics and this is the single major reason for rejection (11, 12). Hence, reliable prediction and modeling of ADME properties is a major concern (13, 14). Obvious as it seems, this realization, however, took considerable time in coming. Nevertheless, it is now widely recognized that physicochemical, pharmacokinetic and biopharmaceutical properties have to be addressed early in drug discovery, an approach that has been tentatively designated as property-based drug design (14).

It is important to realize that the possible (or "allowable") chemical space is incomprehensibly large. Even if

one restricts ourselves to stable and reasonably small compounds (MW < 500) that contain only building blocks common in medicinal chemistry (C, H, N, Of, S, P, F, Cl and Br), the number of possibilities is astronomically high: it has been estimated to be around 10^{62}-10^{63} (15). On one hand, this vastness of the chemical space is a major driving force for a computerized approach because it could be hardly addressed by any other means. On the other hand, it also is a main drawback, as handling chemical compound libraries with 10^{7}-10^{10} structures is already a considerable stretch, especially if 3D conformational issues also have to be addressed. The process, which was brought forward by the development of combinatorial chemistry and high throughput screening, clearly revealed our inability to address spatial (3D, configuration) aspects and the inadequacy of our existing descriptor space. Unfortunately, on the basis that having some kind of prediction is still better than having no prediction at all, it also prompted the development of oversimplified, "quick and dirty" methods designed to handle millions of structures in a field already plagued by the lack of rigorous, scientific, physicochemical-based approaches. A quantity vs. quality schism is becoming increasingly apparent in predictive modeling with less reliable methods designed for high throughput at one end and with much more reliable, high quality but time-consuming methods designed to handle small groups of compounds at the other end.

Computer models

As a further step, whole virtual laboratories aimed to mimic and hopefully to replace significant portions of the real experimental tests that involve cell cultures, animals, or even humans are under development. At the extreme end and still far away, is probably what is called a "virtual patient", software that intends to model all the important aspects of a (human) disease (16). Nevertheless, computerized models are possible and have been implemented every step along the way from drug absorption (17) through pharmacokinetic/pharmacodynamic (PK/PD) modeling (18, 19) to clinical trial outcomes. By integrating structure-generating rules specific to what is designated now as retrometabolic drug design (20, 21) with an analogy-based ranking algorithm that uses property and metabolism estimating subroutines, we implemented an expert system for computer-aided retrometabolic drug design. This not only can generate new soft drug (22) and chemical delivery system (23) structures of interest, but also can assist in selecting the most promising new candidates (24, 25).

Fortunately, computational power increased at a surprisingly steady exponential rate (Fig. 2) and this is a major reason for the increasing relevance of computer-aided drug design (CADD). As predicted by the so-called Moore's law, which held true for more than 3 decades by now, the number of transistors that can be integrated into a microchip and, as a result, processor speed too, doubles about every 18 months, and the trend seems to show

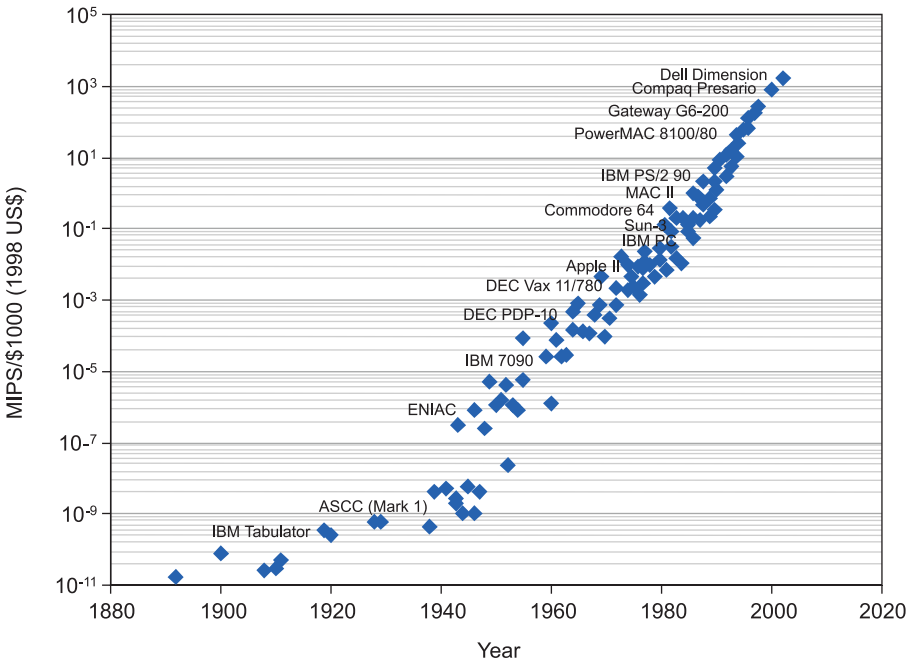

Fig. 2. The development of computational power as measured by the million of instructions per second (MIPS) available per $1000 (1998 US$). The vertical axis that measures computational power has logarithmic scaling; therefore, every major division corresponds to a 100-fold increase. Some representative computers were labeled on their left side. (Graphic prepared using ref. 26).

no indication of slowing down in the near future (26). As a direct result, not only are computerized predictions and models becoming more and more realistic, but sophisticated computations are becoming more and more accessible on inexpensive, mass-produced computers.

Where to start from – descriptors

Certainly, a whole variety of descriptors (parameters) can be used in quantitative analyses (27-34). Good descriptors should be structure-derived, be relevant, have some physicochemical meaning and be easily interpretable by medicinal and/or synthetic chemists. Some of the most frequently employed descriptor classes will be briefly summarized.

Descriptor classes

Size descriptors, such as molecular weight, molecular volume *V*, or surface area *S* are relatively easy to compute; however, accurate and fast volume/surface computations require analytical methods, which may be quite complex (35). Size alone is one of the most relevant parameters for organic liquids (36).

Topological (connectivity) and related parameters (37, 38) or indices, such as those of Kier and Hall (39), Wiener (40), Randic (41), or the more recently introduced electropological indices (42) are also frequently employed mainly because their easy computational availability. They are related to the size and structure of the molecule, but should be used with care because their physicochemical interpretation is rarely clear. It is difficult to imagine what physicochemical relevance or useful medicinal chemical information can be associated with something like the "cube root of the gravitational index", even if it is a descriptor that can be generated from structure and might correlate with certain properties.

Indicator variables are also widely used because they are easy to introduce and calculate and because they are also easy to interpret for synthetic/medicinal chemists. In fact, it is possible to use only indicator variables as structural descriptors. If this is done in a regression analysis approach, the resulting model is the so-called additivity model or Free-Wilson analysis. A major setback is that the use of indicator variables implicitly assumes the additivity of group contributions, which is of course never quite true. That is, it assumes that a group or moiety more or less has the same contribution to the property of interest, which has to be on the right scale, independent of the simultaneous presence of other groups in different molecules.

Electronic descriptors are a very important class. Early parameters, such as Hammett-type polar substituent constant σ (43, 44) and their different variations (σ_m, σ_p, σ^*) or the field and resonance components (F), (R) introduced by Swain and Lupton (45), have been largely replaced by computer calculated, *e.g.*, quantum chemical, descriptors. Many of these, such as dipole moments *D*, polarizabilities α, partial atomic charges q_i, or HOMO-LUMO energies, are certainly relevant and proved useful in numerous quantitative analyses (46). Polarizability and the closely related molar refractivity are electronic parameters, but they are strongly size-related. Combining electronic and structure- or size-related parameters can also provide useful descriptors, and a number of them, *e.g.*, the charged partial surface area, the polar surface area (PSA) (47), or electrostatic potential-modulated surface areas, have been explored. Certain properties, such as pK_a values or NMR chemical shifts δ, are sometimes also used as electronic parameters in predictive approaches and justifiably so. However, these are not structure-derived parameters, but they may be available from experimental determinations or can be relatively reliably estimated.

Steric parameters characterize spatial accessibility. Historically the first, Taft's steric constant E_s (48, 49) was derived based on experimental data. Parameters introduced later were structure-derived and they include, for example, Charton's υ steric parameter (50, 51), STERIMOL parameters *L*, B_1-B_4 and B_5 (52, 53), or simple branching indicators (54, 55). More recently, the inaccessible solid angle Ω_h was also introduced as a measure of steric hindrance (56). Other measures, such as the solvent-accessible surface area of a selected atom or functionality, may also incorporate steric information.

Hydrogen bonding parameters should be very important considering the many roles H-bonds play in aqueous environments because life, as we know it, is water-based. However, they proved somewhat ambiguous to quantify. Both integer only and continuous scales (*e.g.*, fragment contributions or MO calculations) are used. Among the first attempts, Seiler derived group values from differences in octanol-water and cyclohexane-water log partition coefficients (57), Fujita used very simple yes/no indicator variables (58) and Stein and later Lien and coworkers used additivity schemes based on the nitrogen and oxygen atoms present (59, 60). The so-called solvatochromic hydrogen bond acidity and basicity (α^H, β^H), which were introduced within the linear solvation energy relationships (LSER) approach based on solvation effects, are also frequently used. Recently, they were related to fragment constants (61, 62). We also introduced a partition-related *N* descriptor in our QLogP approach (35, 63). A number of other approaches are reviewed in the literature, including more complex, computational model-based H-bond descriptors, such as those in the CoMSIA, GRID, or SuperStar models, or those used in programs such as Hybot, MolSurf, or VolSurf (64-66). Molecular dynamics simulation can also be used to assess H-bond numbers (67). In fact, they provided numbers that could be well modeled by a fragment contribution type approach that also includes consideration of the electronic and steric environment of the H-bonding sites (QikProp model) (68).

Lipophilicity descriptors, such as log *P* or log *k*' chromatographic capacity factors, are among the most

frequently used and most successful parameters. However, they are, in fact, either experimental or calculated/predicted values. The log octanol-water partition coefficient (log $P_{o/w}$) is one of the most frequently employed one and can be relatively reliably calculated based on structure alone (63). It was the subject of many QSPR-type predictive approaches. The abundance of existing methods is well illustrated by the fact that many 4-letter acronyms are already "taken", *e.g.*, ALogP, BLOGP, CLOGP, (ElogP), KLOGP, MLOGP, QLogP, VLOGP, XLOGP. The hydrophobic substituent parameter π, originally introduced (69) by analogy with Hammett-type σ_X constants, can be used as structure-derived descriptor, but one needs not to forget that the actual value of such group contributions is not unequivocal: it strongly depends on the method used for their determination (63).

Seemingly, all these descriptors and/or parameters span a highly multidimensional space. However, because they are strongly intercorrelated, data points corresponding to chemicals of interest are distributed only within a much more restricted subspace. Size- and H-bond-related descriptors seem to be the only ones that consistently surface as relevant for a variety of properties. Early principal component analyses already hinted that for liquids, 2 factors only, one of which is clearly size-related, seem to account for a major part (about 95%) of the variance in the considered properties (70,71). The success of lipophilicity parameters and in particular that of log $P_{o/w}$, may very well be due to the fact that it gives a good representation and a relevant mixture of these 2 descriptors (63).

What is "drug-like"?

Lately, "drug likeness" has become a fashionable term (72). However, one has to remember that it is a fuzzy term: there is no clear delineation between drug and non-drug structures and most likely there never will be one. When dealing with novel structures intended for pharmaceutical use, it is certainly useful to remember that 70% of existing drugs have 0-2 hydrogen bond donors, 2-9 hydrogen bond acceptors, 2-8 rotatable bonds and 1-4 rings (73). It is also useful to remember that certain structural scaffolds are known to be particularly effective for drug design purposes, such building blocks (*e.g.*, benzodiazepines) have been termed "privileged structures" (74). Analysis of the basic ring structure framework of existing drugs reveals surprisingly low diversity: half of the drugs have shapes described by only 32 of the 1179 possible frameworks (75).

A set of heuristic rules known as "Lipinski's rule of 5" (76), which were derived from an analysis of the properties of marketed drugs, has now become almost standard doctrine to avoid permeability and solubility problems and to maximize the chances of surviving development for oral drug candidates. The "rule of 5" requires structures with molecular weight smaller than 500, less than 5 hydrogen bond donors, less than 10 hydrogen bond acceptors and a calculated CLOGP under 5. Furthermore, as medicinal chemists tend to tinker with their original lead structures to improve their properties or to avoid surfacing development problems, comparisons of commercial drugs and their corresponding leads indicated that the original lead structures tend to have lower mass (MW), lower lipophilicity (CLOGP) and less hydrogen bond acceptors than the final drugs (77, 78). On this basis, searches for oral drugs are likely to end successfully if starting from even more restricted libraries (*e.g.*, $100 < MW < 300$, $1 < CLOGP < 3$).

Such observations are certainly useful and should serve as "reality check" in any drug discovery program. Nevertheless, they should not be used as simplistic "hard" filters in managerial-type decision making ("go" *vs.* "no-go"), which is their most frequent current use, but only as "soft" bias in a scientific selection process. A number of existing, top-selling drugs would have not made it through many "drug likeness" filters (79). Furthermore, paradigm changes or shifts may always occur, novel structural motifs may not fit existing rules, drug development techniques are evolving and thinking "out of box" should not be confined to "hard" filtering rules. Because *in silico* and even *in vitro* models are prone to large errors and are still not very predictive of *in vivo* performance, stringent exclusion principles should be combined with liberal promotion principles to identify better overall drugs (80).

How to get there – modeling approaches

Quantitative structure-activity or -property relationship studies attempt to formulate in computerized, mathematically quantified ways the connection between chemical structure and some desired biological (or physicochemical) property and/or activity. More recently, the Hansch school introduced the chem-bioinformatics designation: to understand chemical-biological interactions in mathematical terms (81, 82). From a drug design perspective, the ultimate goal is to describe physiological activity as a function of chemical structure, $A = F(C)$. Because drug distribution and binding processes are equilibrium processes governed by the corresponding free energy differences, $K = e^{-\Delta G/RT} = e^{-(\Delta H - T\Delta S)/RT}$, such relationships should use logarithmic scale and adequate chemical or biological data (*e.g.*, *in vitro* or *in vivo* activities measured by reciprocal molar concentration $1/C$, substrate or receptor binding constants, rate constants, inhibition constants or pharmacokinetic parameters). The logarithmic scales also ensure a normal distribution for the experimental error of biological tests, a requirement for regression-type statistical analyses.

Linear free energy relationship or Hansch approach

A major classic approach commonly used to establish quantitative relationships is the extrathermodynamic

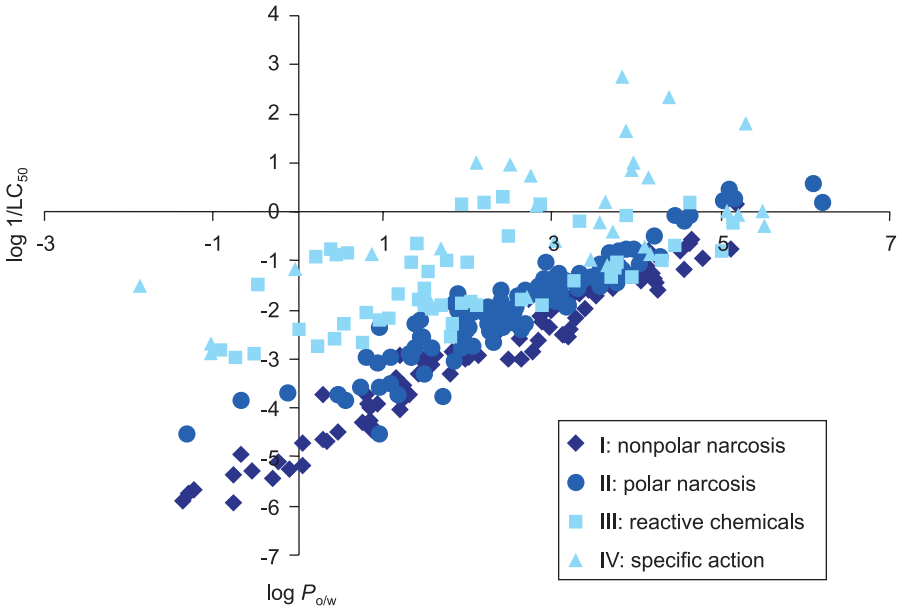

Fig. 3. Aqueous toxicity (measured as the logarithm of the median lethal concentration, log $1/LC_{50}$) data (126) for guppy (*Poecilia reticulata*) as a function of log $P_{o/w}$. If reactive chemicals (aldehydes, epoxides, halophenols, *etc.*) and chemicals with specific action (phosphates, sulphates, pesticides, *etc.*) denoted with light symbols are disregarded, data are distributed around a line showing the excellent predictivity of lipophilicity alone over a range of 8 orders of magnitude.

(linear free energy relationship [LFER] or Hansch) approach (34, 83). This is, in fact, an interproperty relationship that relies on general equations of type:

$$\log 1/C = a_0 + \sum_i a_i p_i \qquad (1)$$

Here p_i are the properties used as descriptors and a_i are coefficients usually determined by linear regression. A typical example would be:

$$\log 1/C = a_0 + a_1 \log P + a_2 (\log P)^2 + a_3 E_s + a_4 \sigma + \ldots \qquad (2)$$

The approach is based on the assumption that having a sufficient number of adequate descriptors (parameters and available properties), the value of other properties/activities of interest can be predicted. Equations of type ***(1)*** also are the basis of most so-called molecular (whole molecule) predictive approaches.

In a geometric interpretation, this requires the data points representing the chemical compounds in the corresponding multidimensional parameter space to be distributed in a subspace of fewer dimensions. This geometric interpretation can be visualized only for the 2D (1 property, 1 descriptor) (Fig. 3) and 3D (1 property, 2 descriptors) case (Fig. 4). Such visualization is also useful in interpreting principal component analysis (PCA) (84), a method frequently used in analyzing large amounts of data. Principal components represent new (mutually orthogonal) axes through the points representing the chemical compounds in the *n*-dimensional space corresponding to the available *n* properties. The first principal component describes the best line through all points and corresponds to the longest dimension. It is some linear combination of the original *n* properties and accounts for the largest possible part of the variance in the data. The second principal component is orthogonal to the first and corresponds to the next longest dimension in orthogonal direction. PCA can be useful if only the first few such principal components can account for most of the existing variance in the data.

Additivity model or Free-Wilson approach

Another major quantitative approach, the Free-Wilson approach (85) (also called additivity model or *de novo* approach) was proposed in an early form by Bruice and coworkers (86) and is used today almost exclusively as modified by Fujita and Ban (87) to correlate biological activity and novel, specific substituent group contributions (a_i)

$$\log 1/C = a_0 + \sum_i a_i n_i \qquad (3)$$

Written in this from with n_i being the indicator (counting) variable for fragment *i*, the analogy with equation 1 is obvious. The approach also forms the basis of most fragment or group contribution predictive methods. Most such methods try to develop a more or less coherent philosophy for their fragmentation system to deal with the issue of chemical significance, but these are in essence

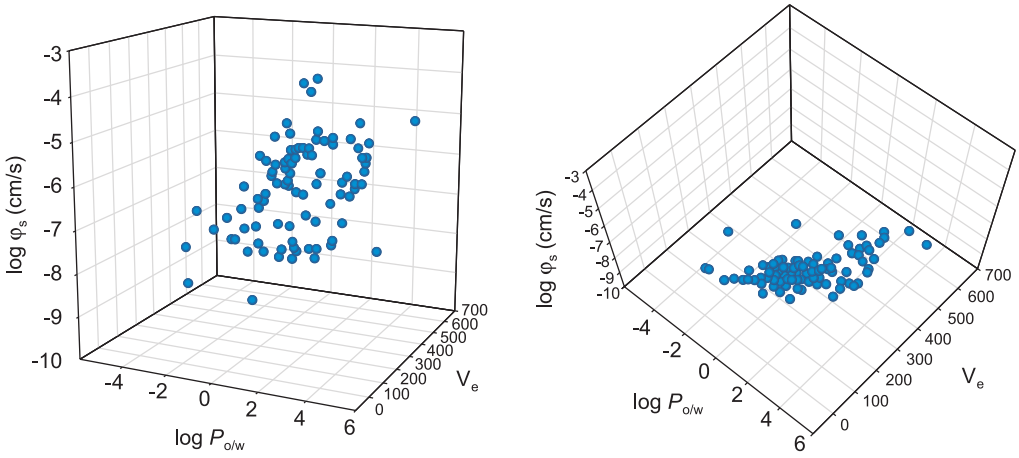

Fig. 4. 3D scatter plot illustrating the dependence of log human skin permeability (logφ_s) on log $P_{o/w}$ and size (measured by V_e) (101). The data are essentially 2D: data points in the 3D space of the corresponding cube are scattered around an inclined planar surface. The left view is from "under" this plane as it rises from the origin in the lower back corner toward the upper, front corner. The right view gives a more tilted view of the same scatter plot, and makes it obvious that most of the 3D space is empty, because points are distributed around a 2D surface.

arbitrary. Using large numbers of different group descriptors may allow a good fit on the training set, but, as with any predictive model, may tend to diminish predictive power outside the original knowledge domain.

Neural network and other approaches

Both previous approaches are essentially regression analysis-type approaches and, therefore, relatively transparent and easy to interpret, a major reason behind their success and popularity. Neural network (NN) approaches represent a possible alternative. They are novel computational techniques that rely on layers of interconnected units in an attempt to model brain function (88) and they were introduced more recently in QSAR and drug design studies (89, 90). NNs can provide excellent nonlinear models, but are difficult to develop and the contribution of input descriptors is problematic to assess ("black box"). In certain cases, multivariate regression models have been shown to outperform NNs by including squares and cross-products of the initial descriptors and by using an orthogonalization procedure to improve descriptor selection (91). Therefore, it remains to be seen whether the possibilities of better predictions and of parsimony in input parameters for NNs will prove sufficient to offset the considerable disadvantages provided by the more work and expertise needed for their development and by the nontransparency of the models as compared to multiple linear regression-type methods.

Many other approaches, which were made possible by the increasing computational power, were explored in the last decade from partial least squares to genetic algorithms or from molecular fingerprints to comparative molecular field analysis (CoMFA). Time will decide which one of them can become truly successful. Many of these approaches can be hampered by the considerable mathematical and programming background required for their implementation and, as a direct consequence, by the possibility that those implementing them often "were 'so intent on making everything numerical' that they frequently missed seeing what was there to be seen" (Barbara McClintock).

Quantitative relationships

QSPR

From a predictive perspective, these should be the easiest to obtain starting from a structure, as "property" in its strictest, more commonly used sense designates only physicochemical properties; hence, no biological systems are involved. Some of the basic physicochemical properties that are of special interest for drug design include, for example, boiling points, melting points, vapor pressure, acidity/basicity (pK_i), lipophilicity (most frequently measured by the log octanol-water partition coefficient log $P_{o/w}$), aqueous solubility, critical micelle concentration, cyclodextrin complexation energies and others. Some of these can be relatively reliably predicted for most compounds of pharmaceutical interest; comprehensive reviews have been published recently (66, 92). Lipophilicity and aqueous solubility received particular interest. Others are more problematic; the melting point is a notorious example because of the complexity of interactions in the crystalline state.

Unfortunately, the field of physicochemical property estimations is highly fragmented with very few attempts made for a unified approach based on the thermodynamics

involved. Besides our unified, molecular size-based approach for nonassociative organic liquids (36, 93), within the pharmaceutical and organic physical chemistry fields, one might mention maybe the attempts by Yalkowsky and coworkers (UPPER) (94) and by Ruelle and coworkers (mobile order thermodynamics) (95).

In a wider sense, QSPR also covers biological properties (96) such as biomembrane permeabilities, including Caco-2 (60, 97), blood-brain barrier (98), intestinal (99), corneal (100) or skin (101) permeability. It can include even complex properties such as intestinal drug absorption/bioavailability (102-107). In fact, in a widest sense, any activity, affinity, or toxicity is a property, but for classification purposes in the biological field, it is probably useful to reserve the QSPR term for the properties that are required to reach the desired target and the QSAR term for the properties that are required to produce the desired activity there (to the extent that they are separable). In the chemical field, especially that of mechanistic organic chemistry, it still remains somewhat arbitrary whether something like reactivity should be considered activity or property.

QSAR

Historically, QSAR is the oldest term. A global search of the Institute for Scientific Information's Web of Science returns 15,157 references for structure-activity, 2783 for structure-property and only 89 for structure-metabolism. If quantitative is also added to the expression, there are only 2496 references for quantitative structure-activity, 294 for quantitative structure-property and only 8 for quantitative structure-metabolism. It is a sobering fact that in all three cases, only about one tenth of the articles deal with quantitative aspects. The origins of the field are usually traced back to a 1962 paper by Hansch, Fujita and coworkers (108), but the terms quantitative structure-activity, -property and -metabolism relationship are first mentioned as such in the ISI database in 1967, 1985 and 1991, respectively.

Already thousands of classical and less classical QSAR have been established (34, 82). Many of them, however, have to be treated cautiously, because they were derived on congener or closely related structures and based on relatively small number of data. Therefore, molecular descriptors might be intercorrelated, the data points to variables ratio might be low and their conclusions might not be generalized to other structures. Despite this and despite the many limitations inherent to any such quantitative model ("All models are wrong, but some are useful." George E.P. Box), they contributed significantly to our present knowledge of relationships between chemical structure and chemical or biological activity (27-30, 32-34). Early filtering ("no-go") decisions cannot be verified, as the corresponding compounds were never made, but they probably resulted in significant time and financial savings by eliminating many nondevelopable compounds (more than likely together with a few active and developable ones as well). In later stages of the drug development process, the "let's measure it anyway" attitude is often still prevailing unless the number of compounds is prohibitive. This, however, is understandable because, as mentioned, most *in silico* and even *in vitro* models are error-prone and poor predictors of *in vivo* performance.

The main contribution of QSAR-type approaches is not to extrapolate and pinpoint highly active new structures, which they are highly unlikely to do, but to slowly transform the field of medicinal chemistry into a rigorous science. Even if often formulated only in hindsight, the ability to attach a quantitative measure to a model is a significant step. Here, at the interface between chemical structure and biological activity, this is often overlooked despite long-standing knowledge formulated by many from Roger Bacon: "If in other sciences we should arrive at certainty without doubt and truth without error, it behooves us to place the foundations of knowledge in mathematics" through Galileo Galilei: "The universe... is written in the language of mathematics" to Eugene P. Wigner: "The language of mathematics reveals itself (to be) unreasonable effective in the natural sciences... a wonderful gift, which we neither understand nor deserve."

QSMR

As mentioned, the importance of early integration of metabolism, pharmacokinetic and general physicochemical considerations in the drug design process has been recognized only relatively recently in industrial settings (13, 14) despite numerous publications describing these concepts and methodologies in the early 1980s (109-111). Hence, QSMR is also a relative newcomer to the field of quantitative analysis. A number of the latest applications as well as computer systems for prediction of xenobiotic metabolism for have been recently reviewed (112, 113). A QSMR that gave good description ($r^2 = 0.80$) of human blood *in vitro* hydrolysis data for 67 ester-containing compounds belonging to seven different noncongener series of drugs has been recently implemented by using the inaccessible solid angle Ω_h calculated around different atoms as a novel measure of steric hindrance (56). Understandably, for metabolism studies, cytochrome P450 (CYP) and its various isoforms are of special interest and they have been the subject of various approaches including, for example, classical LFER for CYP3A4 (114), CoMFA for CYP2C9 (115) and combined protein and pharmacophore model for CYP2D6 (116, 17).

Pharmacophores, the 3D arrangement of molecular features or fragments that are a necessary (but not sufficient) condition for biological activity and pharmacological specificity, attracted considerable computational effort in various fields. The handling of such 3D structures and the identification of corresponding descriptors or relevant molecular superpositions even for large structural libraries are computationally demanding tasks, but considerable progress was made during the last decade

(118-120). The first steps toward identifying the substrates of transporters such as P-glycoprotein have also been made using pharmacophore models (121, 122) or theoretical calculations (123). Because there is increasing evidence that active transport plays just as important a role on the fate of chemicals in living organisms as does metabolism, quantitative structure-transport relationships could become just as important as QSMR.

Conclusions

Tremendous progress was made in the CADD field and such approaches, including QSPR/QSAR/QSMR models, have become essential, indispensable drug design tools. Nevertheless, computerized predictive models, except for a very few cases, are still either not reliable enough or not general enough to be truly useful (and not just in hindsight) as standalone applications. However, if they are well integrated in the overall, collaborative drug design and development process, they can provide, among others, new, even unexpected structures, considerable cost savings, or tools for the understanding, interpretation and visualization of the underlying processes.

References

1. Shilingford, C.A., Vose, C.W. *Effective decision-making: progressing compounds through clinical development.* Drug Discov Today 2001, 6: 941-6.

2. Agnew, B. *When pharma merges, R&D is the dowry.* Science 2000, 287: 1952-3.

3. Aldridge, S. Magic Molecules: How Drugs Work. Cambridge University Press: Cambridge 1998.

4. Giorgianni, S.J. (Ed.) *The pharmaceutical industry at the start of a new century: Unprecedented opportunity, unparalleled challenges.* Pfizer J 1997, 1: 1-37.

5. Le Fanu, J. The Rise and Fall of Modern Medicine. Carrol & Graf: New York 1999.

6. Drews, J. In Quest of Tomorrow's Medicines. Springer: Berlin 1999.

7. Wolff, M.E. (Ed.) Burger's Medicinal Chemistry and Drug Discovery, 5th Ed. Wiley-Interscience: New York 1995.

8. Kuntz, I.D. *Structure-based strategies for drug design and discovery.* Science 1992, 257: 1078-82.

9. Bugg, C.E., Carson, W.M., Montgomery, J.A. *Drugs by design.* Sci Am 1993, 269: 92-8.

10. Joseph-McCarthy, D. *An overview of in silico design and screening: toward efficient drug discovery.* Curr Drug Discov 2002, March 20-23.

11. Prentis, R.A., Lis, Y., Walker, S.R. *Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964-1985).* Br J Clin Pharmacol 1988, 25: 387-96.

12. Brennan, M.B. *Drug discovery. Filtering out failures early in the game.* Chem Eng News 2000, 78: 63-73.

13. Lin, J.H., Lu, A.Y. *Role of pharmacokinetics and metabolism in drug discovery and development.* Pharmacol Rev 1997, 49: 403-49.

14. van de Waterbeemd, H., Smith, D.A., Beaumont, K., Walker, D.K. *Property-based design: Optimization of drug absorption and pharmacokinetics.* J Med Chem 2001, 44: 1313-33.

15. Bohacek, R.S., McMartin, C., Guida, W.C. *The art and practice of structure-based drug design: a molecular modeling perspective.* Med Res Rev 1996, 16: 3-50.

16. Moukheiber, Z. *The virtual patient.* Forbes 2001, January 8: 234-6.

17. Agoram, B., Woltosz, W.S., Bolger, M.B. *Predicting the impact of physiological and biochemical processes on oral drug bioavailability.* Adv Drug Deliv Rev 2001, 50(Suppl. 1): S41-67.

18. Meibohm, B., Derendorf, H. *Basic concepts of pharmacokinetic/pharmacodynamic (PK/PD) modelling.* Int J Clin Pharmacol Ther 1997, 35: 401-13.

19. Meibohm, B., Derendorf, H. *Pharmacokinetic/pharmacodynamic studies in drug product development.* J Pharm Sci 2002, 91: 18-31.

20. Bodor, N., Buchwald, P. *Drug targeting via retrometabolic approaches.* Pharmacol Ther 1997, 76: 1-27.

21. Bodor, N., Buchwald, P. *Retrometabolism-based drug design and targeting.* In: Burger's Medicinal Chemistry, 6th Ed. Abraham, D. (Ed.). Wiley: New York 2002, in press.

22. Bodor, N., Buchwald, P. *Soft drug design: General principles and recent applications.* Med Res Rev 2000, 20: 58-101.

23. Bodor, N., Buchwald, P. *Recent advances in the brain targeting of neuropharmaceuticals by chemical delivery systems.* Adv Drug Deliv Rev 1999, 36: 229-54.

24. Bodor, N., Buchwald, P., Huang, M-J. *The role of computational techniques in retrometabolic drug design strategies.* In: Computational Molecular Biology. Leszczynski, J. (Ed.). Elsevier: Amsterdam 1999, 569-618.

25. Buchwald, P., Bodor, N. *Structure-based estimation of enzymatic hydrolysis rates and its application in computer-aided retrometabolic drug design.* Pharmazie 2000, 55: 210-7.

26. Moravec, H.P. Robot: Mere Machine to Transcendent Mind. Oxford University Press: New York 1999.

27. Higuchi, T., Davis, S.S. *Thermodynamic analysis of structure-activity relationships of drugs: Prediction of optimal structure.* J Pharm Sci 1970, 59: 1376-83.

28. Ariëns, E.J. (Ed.) Drug Design, Vol. 1-10. Academic Press: New York 1971-1980.

29. Purcell, W.P., Bass, G.E., Clayton, J.M. Strategy to Drug Design: A Guide to Biological Activity. Wiley-Interscience: New York 1973.

30. Martin, Y.C. Quantitative Drug Design. A Critical Introduction. Marcel Dekker: New York 1978.

31. Hansch, C., Leo, A.J. Substituent Constants for Correlation Analysis in Chemistry and Biology. Wiley-Interscience: New York 1979.

32. Ramsden, C.A.E. Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study and Therapeutic Application of Chemical Compounds. Quantitative Drug Design, Vol 4. Pergamon Press: Oxford 1990.

33. Hansch, C., Leo, A., Hoekman, D. Exploring QSAR. Hydrophobic, Electronic, and Steric Constants, Vol 2. American Chemical Society: Washington, DC 1995.

34. Kubinyi, H. *The quantitative analysis of structure-activity relationships.* In: Burger's Medicinal Chemistry and Drug Discovery, Vol. 1. Wolff, M.E. (Ed.). Wiley-Interscience: New York 1995, 497-571.

35. Bodor, N., Buchwald, P. *Molecular size based approach to estimate partition properties for organic solutes.* J Phys Chem B 1997, 101: 3404-12.

36. Buchwald, P., Bodor, N. *Simple model for nonassociative organic liquids and water.* J Am Chem Soc 2000, 122: 10671-9.

37. Balaban, A.T., Motoc, I., Bonchev, D., Mekenyan, O. *Topological indices for structure-activity correlations*. In: Steric Effects in Drug Design. Charton, M., Motoc, I. (Eds.). Springer: Berlin 1983, 21-55.

38. Kier, L.B., Hall, L.H. Molecular Connectivity in Structure-Activity Analysis. Wiley: New York 1986.

39. Kier, L.B., Hall, L.H. Molecular Connectivity in Chemistry and Drug Research. Academic Press: New York 1976.

40. Wiener, H. *Structural determination of paraffin boiling points.* J Am Chem Soc 1947, 69: 17-20.

41. Randic, M. *On characterization of molecular branching.* J Am Chem Soc 1975, 97: 6609-15.

42. Kier, L.B., Hall, L.H. Molecular Structure Description. The Electrotopological State. Academic Press: San Diego 1999.

43. Hammett, L.P. *The effect of structure upon the reactions of organic compounds. Benzene derivatives.* J Am Chem Soc 1937, 59: 96-103.

44. Hammett, L.P. Physical Organic Chemistry. McGraw-Hill: New York 1940.

45. Swain, C.G., Lupton, E.C., Jr. *Field and resonance components of substituent effects.* J Am Chem Soc 1968, 90: 4328-37.

46. Karelson, M., Lobanov, V.S., Katritzky, A.R. *Quantum-chemical descriptors in QSAR/QSPR studies.* Chem Rev 1996, 96: 1027-43.

47. Palm, K., Luthman, K., Ungell, A-L. et al. *Evaluation of dynamic polar molecular surface area as predictor of drug absorption: Comparison with other computational and experimental predictors.* J Med Chem 1998, 41: 5382-92.

48. Taft, R.W. Jr. *Polar and steric substituent constants for aliphatic and o-benzoate groups from rates of esterification and hydrolysis of esters.* J Am Chem Soc 1952, 74: 3120-8.

49. Taft, R.W. Jr. *Separation of polar, steric, and resonance effects in reactivity*. In: Steric Effects in Organic Chemistry. Newman, M.S. (Ed.). Wiley: New York 1956, 556-675.

50. Charton, M. *The nature of the ortho effect. II. Composition of the Taft steric parameters.* J Am Chem Soc 1969, 91: 615-8.

51. Charton, M. *The upsilon steric parameter - definition and determination.* Top Curr Chem 1983, 114: 57-91.

52. Verloop, A., Hoogenstraaten, W., Tipker, J. *Development and application of new steric substituent parameters in drug design.* In: Drug Design, Vol. 7. Ariëns, E.J. (Ed.). Academic Press: New York 1976, 165-207.

53. Verloop, A. The STERIMOL Approach to Drug Design. Marcel Dekker: New York 1987.

54. Charton, M. *Steric effects. 13. Composition of the steric parameter as a function of alkyl branching.* J Org Chem 1978, 43: 3995-4001.

55. Austel, V., Kutter, E., Kalbfleisch, W. *A new easily accessible steric parameter for structure-activity relationships.* Arzneim-Forsch 1979, 1979: 585-7.

56. Buchwald, P., Bodor, N. *Quantitative structure-metabolism relationships: Steric and nonsteric effects in the enzymatic hydrolysis of noncongener carboxylic esters.* J Med Chem 1999, 42: 5160-8.

57. Seiler, P. *Interconversion of lipophilicities from hydrocarbon/water systems into the octanol/water system.* Eur J Med Chem 1974, 9: 473-9.

58. Fujita, T., Nishioka, T., Nakajima, M. *Hydrogen-bonding parameter and its significance in quantitative structure-activity studies.* J Med Chem 1977, 20: 1071-81.

59. Stein, W.D. The Movement of Molecules Across Cell Membranes. Academic Press: New York 1967.

60. Ren, S., Lien, E.J. *Caco-2 cell permeability vs human gastrointestinal absorption: QSPR analysis.* Prog Drug Res 2000, 54: 1-23.

61. Platts, J.A., Butina, D., Abraham, M.H., Hersey, A. *Estimation of molecular linear free energy relation descriptors using a group contribution approach.* J Chem Inf Comput Sci 1999, 39: 835-45.

62. Abraham, M.H., Platts, J.A. *Hydrogen bond structural group constants.* J Org Chem 2001, 66: 3484-91.

63. Buchwald, P., Bodor, N. *Octanol-water partition: Searching for predictive models.* Curr Med Chem 1998, 5: 353-80.

64. Dearden, J.C., Ghafourian, T. *Hydrogen bonding parameters for QSAR: Comparison of indicator variables, hydrogen bond counts, molecular orbital and other parameters.* J Chem Inf Comput Sci 1999, 39: 231-5.

65. Böhm, M., Klebe, G. *Development of new hydrogen-bond descriptors and their application to comparative molecular field analyses.* J Med Chem 2002, 45: 1585-97.

66. Matter, H., Baringhaus, K-H., Naumann, T., Klabunde, T., Pirard, B. *Computational approaches towards the rational design of drug-like compound libraries.* Comb Chem High Throughput Screen 2001, 4: 453-75.

67. Duffy, E.M., Jorgensen, W.L. *Prediction of properties from simulations: Free energies of solvation in hexadecane, octanol, and water.* J Am Chem Soc 2000, 122: 2878-88.

68. Jorgensen, W.L., Duffy, E.M. *Prediction of drug solubility from structure.* Adv Drug Deliv Rev 2002, 54: 355-66.

69. Fujita, T., Iwasa, J., Hansch, C. *A new substituent constant, pi, derived from partition coefficients.* J Am Chem Soc 1964, 86: 5175-80.

70. Dunn, W.J. III., Wold, S. *Statistical analysis of the partition coefficient.* Acta Chem Scand B 1978, 32: 536-42.

71. Cramer, R.D. III. *BC(DEF) parameters. 1. The intrinsic dimensionality of intermolecular interactions in the liquid state.* J Am Chem Soc 1980, 102: 1837-49.

72. Ajay, A., Walters, W.P., Murcko, M.A. *Can we learn to distinguish between "drug-like" and "nondrug-like" molecules?* J Med Chem 1998, 41: 3314-24.

73. Oprea, T.I. *Property distribution of drug-related chemical databases.* J Comput Aided Mol Des 2000, 14: 251-64.

74. Fecik, R.A., Frank, K.E., Gentry, E.J., Menon, S.R., Mitscher, L.A., Telikepalli, H. *The search for orally active medications through combinatorial chemistry.* Med Res Rev 1998, 18: 149-85.

75. Bemis, G.W., Murcko, M.A. *The properties of known drugs. 1. Molecular frameworks.* J Med Chem 1996, 39: 2887-93.

76. Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development setting.* Adv Drug Deliv Rev 1997, 23: 3-25.

77. Hann, M.M., Leach, A.R., Harper, G. *Molecular complexity and its impact on the probability of finding leads for drug discovery.* J Chem Inf Comput Sci 2001, 41: 856-64.

78. Oprea, T.I., Davis, A.M., Teague, S.J., Leeson, P.D. *Is there a difference between leads and drugs? A historical perspective.* J Chem Inf Comput Sci 2001, 41: 1308-15.

79. Agrafiotis, D.K., Lobanov, V.S., Salemme, F.R. *Combinatorial informatics in the post-genomic era.* Nat Rev Drug Discov 2002, 1: 337-46.

80. DeWitte, R.S. *On experimental design in drug discovery.* Curr Drug Discov 2002, February 19-22.

81. Hansch, C., Kurup, A., Garg, R., Gao, H. *Chem-bioinformatics and QSAR: A review of QSAR lacking positive hydrophobic terms.* Chem Rev 2001, 101: 619-72.

82. Hansch, C., Hoekman, D., Leo, A., Weininger, D., Selassie, C.D. *Chem-bioinformatics: comparative QSAR at the interface between chemistry and biology.* Chem Rev 2002, 102: 783-812.

83. Hansch, C. *Quantitative structure-activity relationships in drug design.* In: Drug Design, Vol. 1. Ariëns, E.J. (Ed.). Academic Press: New York 1971, 271-342.

84. Jolliffe, I.T. Principal Component Analysis. Springer: New York 1986.

85. Free, S.M. Jr., Wilson, J.W. *A mathematical contribution to structure-activity studies.* J Med Chem 1964, 7: 395-9.

86. Bruice, T.C., Kharasch, N., Winzler, R.J. *A correlation of thyroxine-like activity and chemical structure.* Arch Biochem Biophys 1956, 62: 305-17.

87. Fujita, T., Ban, T. *Structure-activity study of phenetylamines as substrates of biosynthetic enzymes of sympathetic transmitters.* J Med Chem 1971, 14: 148-52.

88. Hinton, G.E. *How neural networks learn from experience.* Sci Am 1992, 267: 144-51.

89. Devillers, J.E. Neural Networks in QSAR and Drug Design. Academic Press: London 1996.

90. Kovesdi, I., Dominguez-Rodriguez, M.F., Orfi, L. et al. *Application of neural networks in structure-activity relationships.* Med Res Rev 1999, 19: 249-69.

91. Lucic, B., Trinajstic, N. *Multivariate regression outperforms several robust architectures of neural networks in QSAR modeling.* J Chem Inf Comput Sci 1999, 39: 121-32.

92. Katritzky, A.R., Maran, U., Lobanov, V.S., Karelson, M. *Structurally diverse quantitative structure-property relationship correlations of technologically relevant physical properties.* J Chem Inf Comput Sci 2000, 40: 1-18.

93. Buchwald, P. *Modeling liquid properties, solvation, and hydrophobicity: a molecular size-based perspective.* Perspect Drug Discov Des 2000, 19: 19-45.

94. Jain, N., Yalkowsky, S.H. *UPPER III: Unified physical property estimation relationships. Application to non-hydrogen bonding aromatic compounds.* J Pharm Sci 1999, 88: 852-60.

95. Ruelle, P. *Universal model based on the mobile order and disorder theory for predicting lipophilicity and partition coefficients in all mutually immiscible two-phase liquid systems* J Chem Inf Comput Sci 2000, 40: 681-700.

96. Blake, J.F. *Chemoinformatics - predicting the physicochemical properties of 'drug-like' molecules.* Curr Opin Biotechnol 2000, 11: 104-7.

97. Camenisch, G., Alsenz, J., van de Waterbeemd, H., Folkers, G. *Estimation of permeability by passive diffusion through Caco-2 cell monolayers using the drugs' lipophilicity and molecular weight.* Eur J Pharm Sci 1998, 6: 317-24.

98. Clark, D.E. *Prediction of intestinal absorption and blood-brain barrier penetration by computational methods.* Comb Chem High Throughput Screen 2001, 4: 477-96.

99. Egan, W.J., Lauri, G. *Prediction of intestinal permeability.* Adv Drug Deliv Rev 2002, 54: 273-89.

100. Edward, A., Prausnitz, M.R. *Predicted permeability of the cornea to topical drugs* Pharm Res 2001, 18: 1497-508.

101. Buchwald, P., Bodor, N. *A simple, predictive, structure-based skin permeability model.* J Pharm Pharmacol 2001, 53: 1087-98.

102. Oprea, T.I., Gottfries, J. *Toward minimalistic modeling of oral drug absorption.* J Mol Graph Model 1999, 17: 261-74.

103. Egan, W.J., Merz, K.M. Jr., Baldwin, J.J. *Prediction of drug absorption using multivariate statistics.* J Med Chem 2000, 43: 3867-77.

104. Wessel, M.D., Jurs, C.P., Tolan, J.W., Muskal, S.M. *Prediction of human intestinal absorption of drug compounds from molecular structure.* J Chem Inf Comput Sci 1998, 38: 726-35.

105. Andrews, C.W., Bennett, L., Yu, L.X. *Predicting human oral bioavailability of a compound: Development of a novel quantitative structure-bioavailability relationship.* Pharm Res 2000, 17: 639-44.

106. Yoshida, F., Topliss, J.G. *QSAR model for drug human oral bioavailability.* J Med Chem 2000, 43: 2575-85.

107. Zhao, Y.H., Le, J., Abraham, M.H. et al. *Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure-activity relationship (QSAR) with the Abraham descriptors.* J Pharm Sci 2001, 90: 749-84.

108. Hansch, C., Maloney, P.P., Fujita, T., Muir, R.M. *Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients.* Nature 1962, 194: 178-80.

109. Bodor, N. *Novel approaches for the design of membrane transport properties of drugs.* In: Design of Biopharmaceutical Properties through Prodrugs and Analogs. Roche, E.B. (Ed.). Academy of Pharmaceutical Sciences: Washington, DC 1977, 98-135.

110. Bodor, N. *Designing safer drugs based on the soft drug approach.* Trends Pharmacol Sci 1982, 3: 53-6.

111. Bodor, N. *Soft drugs: Principles and methods for the design of safe drugs.* Med Res Rev 1984, 3: 449-69.

112. Wessel, M.D., Mente, S. *ADME by computer.* Annu Rep Med Chem 2001, 36: 257-66.

113. Langowski, J., Long, A. *Computer systems for the prediction of xenobiotic metabolism.* Adv Drug Deliv Rev 2002, 54: 407-15.

114. Riley, R.J., Parker, A.J., Trigg, S., Manners, C.N. *Development of a generalized, quantitative physicochemical model of CYP3A4 inhibition for use in early drug discovery.* Pharm Res 2001, 18: 652-5.

115. Rao, S., Aoyama, R., Schrag, M., Trager, W.F., Rettie, A., Jones, J.P. *A refined 3-dimensional QSAR of cytochrome P450 2C9: Computational predictions of drug interactions.* J Med Chem 2000, 43: 2789-96.

116. de Groot, M.J., Ackland, M.J., Horne, V.A., Alex, A.A., Jones, B.C. *Novel approach to predicting P450-mediated drug metabolism: development of a combined protein and pharmacophore model for CYP2D6.* J Med Chem 1999, 42: 1515-24.

117. de Groot, M.J., Ackland, M.J., Horne, V.A., Alex, A.A., Jones, B.C. *A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed N-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6.* J Med Chem 1999, 42: 4062-70.

118. Miller, M.D., Sheridan, R.P., Kearsley, S.K. *SQ: A program for rapidly producing pharmacophorically relevent molecular superpositions.* J Med Chem 1999, 42: 1505-14.

119. Mason, J.S., Good, A.C., Martin, E.J. *3-D pharmacophores in drug discovery.* Curr Pharm Des 2001, 7: 567-97.

120. Tropsha, A., Zheng, W. *Identification of the descriptor pharmacophores using variable selection QSAR: Applications to database mining.* Curr Pharm Des 2001, 7: 599-612.

121. Seelig, A. *A general pattern for substrate recognition by P-glycoprotein.* Eur J Biochem 1998, 251: 252-61.

122. Penzotti, J.E., Lamb, M.L., Evensen, E., Grootenhuis, P.D. *A computational ensemble pharmacophore model for identifying substrates of P-glycoprotein.* J Med Chem 2002, 45: 1737-40.

123. Österberg, T., Norinder, U. *Theoretical calculation and prediction of P-glycoprotein-interacting drugs using MolSurf parametrization and PLS statistics.* Eur J Pharm Sci 2000, 10: 295-303.

124. Reuben, B.G. *The consumption and production of pharmaceuticals.* In: The Practice of Medicinal Chemistry. Wermuth, C.G. (Ed.). Academic Press: London 1996, 903-38.

125. Pharmaceutical Research and Manufacturers of America 2001 Industry Profile. PhRMA: Washington DC, 2001.

126. Katritzky, A.R., Tatham, D.B., Maran, U. *Theoretical descriptors for the correlation of aquatic toxicity of environmental pollutants by quantitative structure-toxicity relationships.* J Chem Inf Comput Sci 2001, 41: 1162-76.