Fragmental Methods in the Analysis of Biological Activities of Diverse Compound Sets

P. Japertas[1,2], R. Didziapetris[1] and A. Petrauskas[1*]

[1]*Pharma Algorithms, Inc., Tauro 12, Vilnius 2001, Lithuania*

[2]*Vilnius University, Faculty of Chemistry, Naugarduko 24, Vilnius 2006, Lithuania*

Abstract: The current mini-review explains how fragmental methods (FMs) can be used in the analysis and prediction of physicochemical properties and biological activities. The considered properties include log *P*, solubility, pK_a, intestinal permeability, P-gp substrate specificity and toxicity. The focus will be a description of a "mechanistic" approach, which implies a gradual reduction of alternative explanations for any property or activity. This means a flexible construction of fragmental parameters using large amounts of experimental data. Since biological activities involve multiple (unknown) target macromolecules with multiple binding modes, a stepwise classification (C-SAR) analysis is most useful. It involves the following procedures: (i) construction of physicochemical profiles using parameters that can be reliably predicted, (ii) identification of reactive functional groups and the largest active skeletons, (iii) generalization of these groups and skeletons in terms of "site-specific physicochemical profiling". This entails a dynamic construction of 2D pharmacophores that can be converted into 3D models.

Keywords: Fragmental methods, biological activity prediction, mechanistic approach, classification analysis.

1. INTRODUCTION

The analysis of biological activities (BAs) aims at the design of new compounds using a wide variety of methods [1-28]. Selecting the particular method depends on the complexity of the biological system, the amount and reliability of experimental measurements, and on the diversity of compounds. Several decades ago, the design of new compounds was based mostly on a chemist's intuition that was reinforced by QSAR analyses of small congeneric sets [1-5]. In the 1990s, structure-based design using 3D QSAR methods became popular [7-9]. These methods help our intuition by providing a means to visualize ligand-receptor interactions, but they cannot be used in the analysis of diverse compounds due to problems of molecular alignment [7,8]. Today attention is being shifted to the analysis of large sets of diverse compounds using various machine learning [10,11], similarity searching [12-14], "global QSAR" [15-19] and classification (C-SAR) [20-24] methods. Most of these methods have appeared as a response by computational chemists to the explosion of chemical and biological information generated by rapid advances in combinatorial chemistry and HTS. The original idea behind congeneric QSARs was to test scientific hypotheses [2-5,25-27]. The new generation of similarity and "global" QSARs aims at answering a pragmatic question – which compounds should be synthesized? This indicates a potentially big problem, as the synthesis of new compounds using sophisticated similarity and diversity analyses does not necessarily lead to rational drug discovery [9]. Given the fact that large amounts of experimental data for diverse compounds have already been (and continue to be) generated, how can we use this data more effectively in the design of new compounds? The question comes down to the effective extraction of new mechanistic information from the existing data sets. Recently we attempted to provide a solution for this task by developing a new software system, Algorithm Builder (AB) [28,29], using a variety of fragmental methods (FMs). The present mini-review is largely based on the thoughts and experiences gathered during the continuous development of AB and related predictive algorithms [30-34].

2. MECHANISTIC ANALYSIS

Extracting new mechanistic information from any data set depends on our ability to use the existing information about the related chemical and biological mechanisms [2,4]. Several reviews [3-5,25-27] provided excellent examples of how this can be done in the analysis of congeneric data sets. In the case of diverse compounds, the extraction of new mechanistic information is a much more difficult task due to the almost infinite complexity and fuzziness of biological constituents [25-27,35]. This leads to the extreme sensitivity of the obtained results in relation to the clarity of descriptors [28] and logical sequence of explanations [32].

2.1. Descriptors

All descriptors can be subdivided into three groups, depending on the clarity of chemical and biological explanations, as shown in Table 1.

The first group involves physicochemical parameters, such as pK_a, solubility, log *P*, H-bonding (solvation parameters), and some other descriptors. These are most useful in the analysis of "non-specific" biological effects that are not caused by reactive functional groups or active skeletons. The second group involves "mechanistic

*Address correspondence to this author at the Pharma Algorithms, Inc., Tauro 12, Vilnius 2001, Lithuania; E-mail: Alanas@ap-algorithms.com

1389-5575/03 $41.00+.00

Table 1. Rough Estimation of the Usefulness of Various Descriptors in the Mechanistic Analysis of Complex BAs.

	Type	Descriptors	Interpretations
I.	Physchem	*MV*, *MW*, *MR*, T*PSA*, Abraham's H-bonding (α and β)	Simple
		Log *P*, solubility, ionization (pK_a)	Simple[a)]
II.	"Mechanistic fragmental"	Functional groups and pairs of groups (interactions)	Moderate
		Large chains and HOSE codes	Moderate
		Rigid rings (scaffolds)	Moderate
III.	"Statistical fragmental"	Small chains, HOSE codes or multilevel neighborhoods of atoms	*Difficult*
		Atoms or e-state indices	*Difficult*
		Fingerprints, hashkeys	*Difficult*
		Atom pairs, triplets & quadruplets	*Difficult*
IV.	Other (see article by Gasteiger in this issue)	Topological, shape, geometrical, Quantum chemical	*Difficult*

[a)] Accurate calculation of these descriptors is problematic (see article by Caron co-workers in this issue); one has to be very careful when using them in the analysis of BAs.

fragmental" parameters, such as conventional functional groups, interactions (pairs of groups), large atom chains and rigid rings (scaffolds). These are useful in describing "specific" chemical and biological effects, including the following: (i) intra-molecular interactions, (ii) ionization and chemical reactivity of small functional groups, (iii) "pharmacophoric" effects of large skeletons. All of these parameters are obtained by certain rules that split compounds into constituent parts. Each constituent part represents an integer variable, denoting the number of occurrences of a given fragment (or ensemble of fragments) in a compound. Any given parameter is only useful in mechanistic explanations if it preserves the integrity of a given functional group or active skeleton. Not all FMs can satisfy this condition. "Statistical fragments" split reactive groups and active skeletons into smaller parts, and thus resemble many "other" computational descriptors with ambiguous explanations. All of these descriptors will receive only minor consideration in this mini-review.

2.2. Multi-Step Analysis

Frequently "mechanistic analysis" is understood as the derivation of any QSAR with simple-to-understand descriptors. Such a definition is far from complete when considering diverse compounds that may cause a great variety of chemical and biological effects [25-27]. Simple descriptors may not always provide simple explanations, as the number of possible effects may exceed the number of observations. Multiple possible explanations must be analyzed in a deductive way using the concept of "limiting factor" as proposed by Albert [1,35]. Once the complexity of explanations is reduced to a certain level, we can apply the existing mechanistic information to generate new knowledge. As Hansch and co-workers [27] pointed out, "one can get surprising information out of a very simple system if one can make use of known chemical and biological information". This statement expresses the essence of any stepwise analysis – utilize the existing knowledge by all the possible means, only then seek statistical significance and predictive power. Fig. (**1**) shows this as a three-step scheme (A-C), where statistical significance originates from mechanistic generalizations.

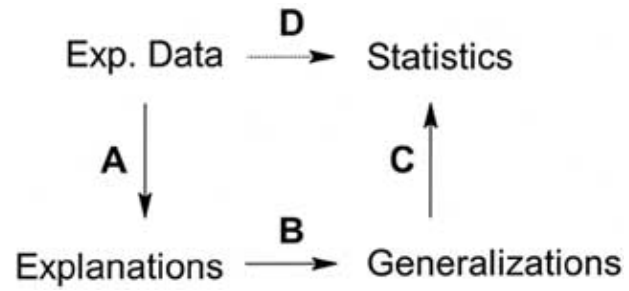

Fig. (1). A – stepwise reduction of multiple explanations, B – classification of effects and construction of new parameters, C – reduction of parameters based on mechanistic generalizations, D – single-step statistical optimization without mechanistic analysis.

The analysis of BAs involves a dynamic construction of new parameters using the following principles: (A) find the simplest possible explanations, (B) convert explanations into parameters, (C) generalize parameters to achieve predictive power. Hansch and Leo [36-38] used this approach in deriving their CLOGP model. Lipnick [35] attempted to use a similar approach in predicting rodent acute toxicity (LD_{50}, although not using FMs explicitly). We [32-34] used this approach in the analysis of LD_{50}, human intestinal absorption (*HIA*) and P-gp substrate specificity.

2.3. Single-Step Analysis

Frequently the mechanistic approach cannot be used due to the lack of sufficient information about the underlying chemical and biological mechanisms. In such cases, a "purely statistical" approach is used. The latter originates from the belief that good statistical validation of any model provides enough information for accepting or disproving any

hypotheses. It can also be associated with the assumption that similar compounds possess similar activities [7,9]. Fig. (**1**) shows this as a single-step approach (D) that achieves good statistical significance without a detailed analysis of the underlying effects. Any "statistical" parameters from Table 1 can be used in this approach, provided they are reinforced by certain statistical methods. From the computational chemist's point of view, most FMs produce too many variables with low statistical significance. Various statistical methods (PCA, PLS and others, see article by Migliavacca in this issue) can resolve this problem by combining the original descriptors into some derivative functions [10,18,19,39]. This results in a loss of explanatory power, as derivative functions do not preserve mechanistic reasoning. The obtained predictive power can be useful in many practical situations, but it cannot be used in a mechanistic analysis of BAs. This was quite clearly stated by Lipnick [35] and Testa [26], who criticized the use of any statistical parameters in QSAR analysis. Such a criticism can be generally applied to all statistical models that use descriptors with ambiguous interpretations and/or employ the concept of structural similarity.

3. QSPR ANALYSIS

QSPR analysis aims at predicting various physicochemical properties that offer simple interpretations in the analysis of BAs. All of such properties can be subdivided into two groups, as shown in Table 1. We will only consider analysis and prediction of log *P*, pK_a and solubility, as most of the remaining parameters can be calculated using simple atomic increments [28,40]. While log *P* can be quite satisfactorily predicted by a variety of FMs [40-42], this is not exactly so in the case of pK_a and solubility [40]. An analysis of these properties is useful in demonstrating the existing limitations of any QSPR predictions.

3.1. Log *P* Prediction

Several books [36,37,43,44] and review papers [28,38,40,41] have already described the use of FMs in log *P* predictions. Most of these methods imply a summation of fragmental increments and correction factors that account for intra-molecular interactions between functional groups. As Hansch and Leo [37] pointed out, an understanding of these interactions can provide additional insight in the design of new drugs. The problem is that there are so many types of internal interactions that they are difficult to generalize. One can mention inductive, resonance, steric, H-bonding, and alpha-effects. Each effect depends on particular functional groups and the molecular skeletons between the groups. The simplest explanations can be achieved using "IC-based" fragmentation [28] as illustrated in Fig. (**2-A**).

By this method, a compound is split into parts using the following steps: (i) defining "isolating carbons" (IC_i) as any carbons that are not doubly or triply bonded to heteroatoms, (ii) defining functional groups (F_j) as any inter-bonded atoms without ICs, (iii) defining interactions (Int_{ijk}) as any pairs of functional groups separated by certain skeletons. As a result, the following equation is obtained:

$$\text{Log X} = \Sigma\, a_i \text{ x } IC_i + \Sigma\, b_j \text{ x } F_j + \Sigma\, c_{ijk} \text{ x } Int_{ijk} \qquad \text{(eq 1)}$$

Here log X is log *P*. The obtained fragments correspond to conventional functional groups that can be analyzed in terms of electron withdrawing capability, ionization and chemical reactivity. For example, $H_2N\text{-}CH_2\text{-}COOH$ is automatically split into H_2N, CH_2 and COOH, whereas $H_2N\text{-}CO\text{-}CH_3$ is split into H_2NCO and CH_3. The obtained interactions (Int_{ijk}) can be related to various types of internal effects. For example, a fragmentation of salicylic acid produces an aromatic chain HO-C-C-COOH that can be associated with strong ortho-interaction (as a combination of internal H-bonding and electronic resonance). A fragmentation of lactic acid produces an aliphatic chain HO-

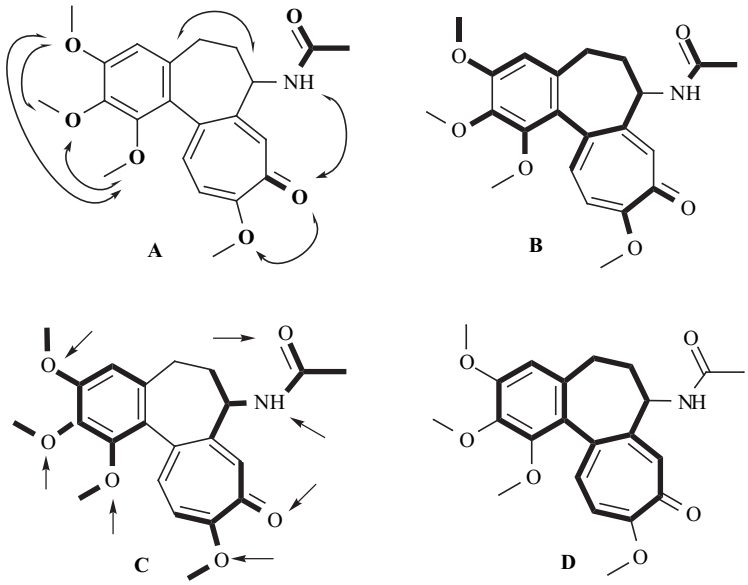

Fig. (2). Structure fragmentations: A – IC-based groups and interactions (IC-atoms not highlighted, arrows denote interactions); B – atom chains (only some of the most meaningful 4-atom chains are shown); C – two-layer HOSE codes (arrows denote polar atom centers); D – rigid skeletons.

C-COOH that can be associated with strong inductive effects. Such physicochemical clarity can only be obtained at the cost of reduced statistical significance. If all physicochemical effects are represented by individual parameters, their number becomes comparable to the number of experimental data points. This leads to many single-point determinations and complex linear dependencies. In order to overcome this problem, Hansch and Leo [36-38] used the "constructionist" approach. This implies a stepwise analysis of compounds with different functionality in a deductive way. The log *P* values of simple hydrocarbons provide increments for IC atoms (these differ in their branching, cyclization and aromaticity). When IC increments are subtracted from the log *P* of mono-functional compounds, increments for individual functional groups are obtained. When these are subtracted from log *P* of bi-functional compounds, increments of interactions are obtained. The analysis continues within specific classes of poly-functionals in order to understand their internal interactions. At each step various types of topological and physicochemical generalization can be made, so the total number of parameters can be quite small. In Fig. (**1**), this approach corresponds to steps A-C, leading to the "mechanistic induction" that is the major source of predictive power in CLOGP. Many subsequent investigators attempted to maximize the statistical significance of structural descriptors using steps B and C. This led to the derivation of new FMs with varying predictive power [40-42]. The KOWWIN method of Meylan and Howard [45] deserves a special mention, as it produced good statistical results on several test sets [40-42]. A common limitation of any FMs that were designed for predicting log *P* is that they do not preserve the mechanistic reasoning of parameters when used for predicting other properties. Different properties require constructing different FMs in keeping with steps A-C. If this is not done, we cannot generate mechanistic induction comparable to that achieved in CLOGP. Recently we [28] proposed a possible solution for this problem using equation (1). All parameters were clustered using an empirical similarity key and hierarchical clustering analysis (HCA). Such a procedure can preserve the physicochemical meaning of parameters while reducing their number to almost any desired level. The problem is that any given similarity key has quite a narrow practical application. The similarity key that was constructed for log *P* may not be applicable to other physicochemical properties.

3.2. Limitations of log *P* Calculations

The predictive power of any method is limited to the knowledge that can be extracted from the training set (see article by Caron co-workers in this issue). The currently available mechanistic knowledge is unsatisfactory in the following areas: (i) strong alpha-interactions in poly-hetero-cycles and (ii) conformational effects in natural compounds. As Leo correctly pointed out [38], "many chemists are understandably skeptical of fragmenting an aromatic hetero-cycle and trying to make sense of the parts". A similar statement can also apply to the fragmenting of large peptides, sugars and alkaloids. Even more serious problems arise when predicting log *P* of electrolytes that involve the partitioning of charged species. The latter effect is difficult to model due to the complexity of ion partitioning [46] and the lack of sufficient experimental data. For all of these reasons, any log *P* calculations for compounds that are very different from the training set may be inaccurate. A highly incorrect assumption is that a lack of mechanistic knowledge can be compensated for by statistical approaches, such as ANNs with e-state indices [39] or PLS with atom chains [19]. These approaches are quite efficient when seeking to obtain predictive power in a hurried manner. But they do not generate any mechanistic induction, so the limits of their practical application can hardly be ascertained. The actual difference between statistical and mechanistic inductions is difficult to estimate. Recently Mannhold and Petrauskas [42] attempted to do this in a comparative validity test of the following programs: (i) CLOGP and KOWWIN - developed using steps A-C in a manual way, (ii) AB/LogP – using steps B-C in an automated way, and (iii) SciLogP Ultra – using step D (ANNs with e-state indices). The obtained results indicate that the accuracy of predictions decreases proportionally to the decrease in the amount of mechanistic interpretations. To generalize this statement, the following order of importance for predictive power can be set forth: training set > mechanistic induction > statistical induction.

3.3. pK_a Prediction

pK_a is perhaps the most important physical property in the analysis of BAs, but it is also the hardest to predict. Frequently a compound's ionization affects BA in a much stronger way than log *P* or any other properties [47]. This raises stringent requirements for the accuracy [48] and speed of pK_a calculations that cannot be met by any computational methods. Da Silva and co-workers [49] attempted predicting pK_a using *ab initio* calculations. Citra [50] used semiemprical quantum mechanics. Klopman and Fercu [51] attempted using simple functional groups. Xing and Glen [52] used small HOSE codes (similar to those used in NMR predictions [53]) with PLS. All of these studies dealt mostly with small mono-electrolytes, and in many cases produced not very accurate calculations even for compounds from the training sets. Meanwhile, the analysis of BAs frequently involves large poly-basic (drug-like and natural) compounds with an increased complexity of electrolyte behavior. How can we predict accurate pK_a values for such compounds? The most reliable predictions are achieved using Hammet or Taft equations based on substituent effects. However, these equations are only valid for narrow congeneric sets, whereas substituent sigma constants depend on ionization center. Perrin and co-workers [54] described many additional pitfalls: (i) tautomerism and vynilogy, (ii) charge transfer in conjugated (hetero-aromatic) systems, (iii) the effect of multiple ionized groups on a given electrolyte center. None of these effects can be accurately predicted in a high-throughput mode, unless the "generic substructure" approach is used. The latter implies an empirical definition of multiple structural skeletons using various types of generic atoms, bonds and functional groups. Each skeleton must be provided with a generalized Hammett-type equation(s) that accounts for the variable substituent effects. The definition of generic substructures can be compared to the lengthy constructionist approach using steps A and B in Fig. (**1**).

Unlike in log *P*, the key structural parameters that affect pK_a cannot be obtained by any automated means. Potentially, this could be achieved by using large HOSE codes (similar to those used by Xing and Glen [52]), but such a suggestion still requires a careful verification. In the meantime, manual analysis remains the only way of achieving any reliable predictions, so it is easy to understand why there are so few commercial programs for predicting pK_a values [30,55-57]. All of these programs can be viewed as different combinations of the generic substructure and Hammett approaches. The validity of these programs depends on the amount of mechanistic interpretation that was applied in their development, whereas the underlying theories are not as important.

3.4. Aqueous Solubility Prediction

The prediction of solubility (log S_W) requires considering a compound's "crystallinity" – a phenomenon that is not well understood. So it cannot be predicted using any meaningful basis, except that we can try various hypotheses. Bearing this in mind, various versions of the following general equation are used [40,45,58-62]:

$$\text{Log } S_W = \text{Const} + \left(\Sigma\, a_i \text{ x } f_i\right) + b \text{ x } mp + c \text{ x log } P + \text{other properties} \quad \text{(eq 2)}$$

Here S_W means characteristic solubility of neutral compounds. The first term denotes a sum of fragmental increments or correction factors [40,45,58,59]. All of these increments are frequently omitted [60-62], as the remaining physicochemical parameters can provide good enough correlations. This observation is only valid when dealing with simple, not-very-diverse compounds. The second term, *mp*, is melting point that is supposed to reflect the strength of crystal packing [60-62]. It is also frequently omitted, as it cannot be accurately calculated [40]. Abraham and Lee [62] proposed to replace *mp* with a product of H-acidity and basicity (α multiplied by β) in an amended solvation equation. This model provides very intelligent explanations, but it includes a polarity-polarizability (π) parameter that is difficult to calculate [63]. In addition, this model was derived using quite a limited data set (N = 659), covering mostly compounds with only one or two functional groups. The third parameter that deserves special consideration is log *P*. Its use can be justified by an empirical correlation between log S_W and log *P* that was obtained by Hansch and co-workers for 156 liquids [64]. Testing this correlation on a larger data set (N = 3,738) using mostly solids revealed that it only holds for compounds with log $P > 0.5$ (N = 3,000, $R^2 = 0.7$, SD = 1.1). Hydrophilic compounds with log $P < 0.5$ disobey this correlation (N = 738, R = 0), indicating that equation (2) becomes invalid. Sometimes log *P* is replaced with log *D* (the "distribution" coefficient for electrolytes at a given pH), enabling the one to estimate the pH dependence for log Sw [58]. This approach deserves serious criticism, as equation (2) is generally not valid for electrolytes. Variable pH and different counter-ions may influence crystal packing – a factor that may nullify any quantitative predictions. So the solubility of electrolytes can hardly be predicted by any QSPR methods, as the underlying LFER principle is easily compromised. One can only seek qualitative predictions by means of classification methods. Fig. (**3**) shows an example of such an approach [31].

Plot (A) displays a subdivision of all the possible electrolytes into several classes. Plot (B) shows a further subdivision of each electrolyte class into smaller sub-classes using log *P* and *MW* cut-off values. Each cut-off value is defined using C-SAR methods. Such "physicochemical profiling" (PP) is the only way to go when the mechanistic factors are not well understood. In this case FMs are not as useful, as we do not know how exactly the obtained parameters should be interpreted.

4. QSAR ANALYSIS

The major (but not the only) goal of accurate physicochemical predictions is to facilitate the analysis and interpretation of BAs. This can be done in two different ways - using QSAR and C-SAR methods. For the sake of simplicity, we will define QSAR as any single equation that relates BAs to some descriptors from Table 1. Such a simplified definition is very convenient, since it distinguishes QSAR from C-SAR that describes BA by a set of multiple equations. Generally, the "single-equation" models are much more popular than the "multiple-equation" models, although in the case of diverse compounds they have numerous limitations.

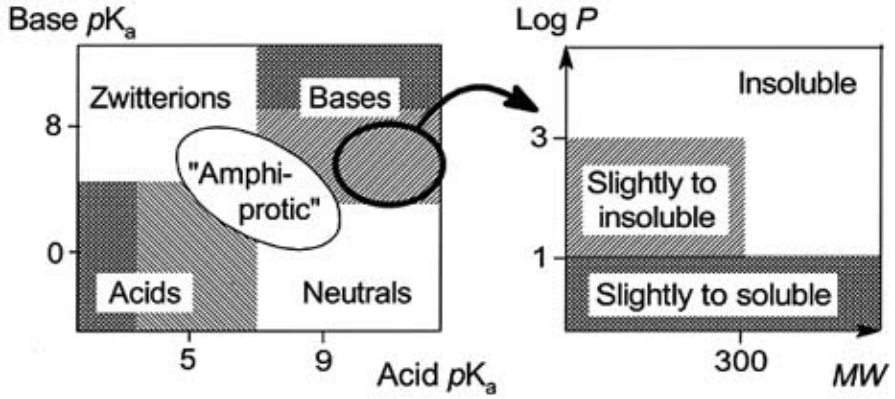

Fig. (3). Physicochemical Profiling (PP) for solubility of crystalline compounds.

4.1. Assumptions and Limitations

The major limitation of any QSAR is that it assumes the existence of LFER. Species with different charges, *MW*, T*PSA* and other properties may alter biological mechanisms [25,26]. So any single equation that attempts to describe a large set of diverse compounds has no mechanistic rationale. This limitation can only be overcome when dealing with small sets of not-very-diverse, non-reactive, non-electrolyte compounds. In addition, these compounds must not involve unknown pharmacophores that may induce various biological targets. This forces us to consider any biological system as "non-specific". The latter assumption is an obvious oversimplification, so a variety of 3D methods have been designed [6-8]. Most of these methods assume that BA is "highly specific", originating from a single static receptor with an invariable binding mode. Such an assumption is another oversimplification [7,8,26], as any *in vivo* (and many *in vitro*) systems include multiple targets with fuzzy specificity. Experimental data from such biological systems cannot be described by any single equation or 3D model. Instead, multiple equations and 3D models must be considered, leading us to the "mechanistic C-SAR" approach. Fig. (**4**) shows how mechanistic information depends on the initial assumptions made by each method. Continuous arrows indicate trends of the mechanistic approach that is also referred to as "baseline analysis". Dotted arrows indicate trends of the statistical or similarity approaches that put statistical considerations in front of mechanistic analysis.

The difference is in how we use the available mechanistic information in the analysis of experimental data. The baseline approach attempts to utilize this information to the maximum extent, whereas the similarity approach mainly disregards it.

4.2. Baseline Analysis

The baseline approach originates from the idea that outliers from QSAR predictions indicate new biological mechanisms. This idea is only valid if the underlying QSAR model reflects a certain biological mechanism by itself, making the analysis of outliers simpler than the analysis of parent BAs. As Kubinyi [5] pointed out, "Predictions from QSAR studies should mainly serve to derive new hypotheses". Derivation of new hypotheses implies using mechanistic deductions that cannot be replaced with statistical inductions [26,32,35]. Lipnick [35] provided an excellent example of the baseline approach in predicting LD_{50} values. The analysis started with the simplest possible compounds – mono-functional alcohols and ketones – that were described in terms of a bi-linear log *P* model. This model reflects an idea of non-specific narcosis that is believed to be the primary cause of acute toxicity of any non-reactive non-electrolytes [65]. The obtained QSAR was used to analyze complex poly-functionals that produced positive deviations from baseline predictions. These were grouped into several classes according to the following factors: (i) "biological similarity" of compounds based on the chemical reactivity of functional groups, and (ii) "excess toxicity" expressed as the difference between the actual and predicted toxicities. Such a classification can be compared to mechanistic C-SAR analysis [32], the major difference being in how we define the "biological similarity" of outliers. In the baseline approach it is defined quite loosely, using the investigator's intuition. In the C-SAR approach it is defined very strictly, using various descriptor cutoff values. The baseline approach can also be compared to the Hansch and Leo constructionist approach in developing CLOGP. Both of these approaches imply a very careful analysis of experimental data using steps A and B in Fig. (**1**). The difference is that any BA is a much more complex property than log *P*. Accordingly, constructing any credible algorithm would require much greater efforts than developing CLOGP. For this reason, there are no examples of mechanistic baseline analysis using FMs. Any FMs produce large numbers of parameters that are difficult to relate to various chemical and biological effects. If any parameters do not receive unambiguous interpretations, the mechanistic analysis of outliers becomes impossible. If all parameters receive clear interpretations, we obtain a "knowledge base" that considers BA far beyond the frameworks of a single QSAR equation.

4.3. Similarity Analysis

Similarity analysis can be defined as any method that does not provide unambiguous interpretations for all QSAR

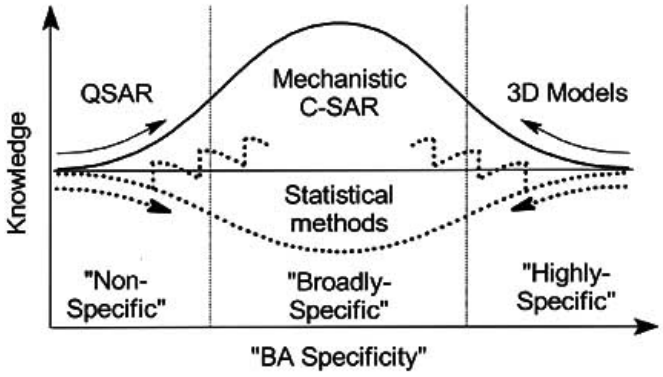

Fig. (4). Application areas of QSAR, 3D and C-SAR methods. Continuous lines denote the mechanistic approach, dotted lines – the statistical approach, vertical dotted lines – new experimental work. Arrows denote trends in multi-step analyses.

(or C-SAR) parameters. This is always the case if in Fig. (**1**) statistical optimization (step C) is not preceded by a careful constructionist approach using steps A and B. So we come down to a single-step statistical analysis (D) that disregards the consideration of the particular biological mechanisms. The only reasoning behind such an analysis is that similar compounds possess similar BAs. The latter assumption is not always correct, as even small changes in chemical structure may lead to dramatic changes in BA [7,9]. Frequently the similarity approach is applied for predicting various BAs using "statistical" fragmental parameters. For example, Andrews and co-workers [15] used molecular fingerprints in predicting oral bioavailability. Ghuloum and co-workers [17] used molecular hashkeys (representing certain combinations of fragmental and non-fragmental descriptors) in predicting a variety of pharmaceutical properties. Anzali and co-workers [18] used "multilevel neighborhoods of atoms" (analogs of small HOSE codes) in predicting over 500 types of BAs. The striking feature of all of these studies is that the obtained models cannot be used for baseline predictions. Any similarity prediction is nothing more than a statistical hypothesis that co-exists with many other hypotheses. New hypotheses (that explain deviations from predictions) cannot be based on the old hypotheses, so the similarity predictions cannot be used for baseline predictions. The dotted line on Fig. (**4**) shows what happens if one attempts to do so. Each step of statistical induction without mechanistic deduction decreases our ability to understand the reality. This can be associated with the risk of error propagation [26], leading to proliferation of unproductive hypotheses. The derivation of any statistical model must be followed by new experimental work (synthesis and screening), which is also unproductive, as it does not aim at testing any rational hypotheses. The result is the generation of large amounts of experimental data that cause "increasing the size of the haystack" when "trying to find a needle" [9].

5. C-SAR ANALYSIS

Classification analysis differs from QSAR in that it is not restricted by the limitations of LFER. Instead of deriving a single QSAR equation, multiple equations (or rather inequalities) are derived. These group compounds into classes that may correspond to different biological mechanisms. The question of whether any given classification is biologically meaningful is of utmost importance. Among a great variety of methods that group compounds into classes [10,11,20-25], we will only consider recursive partitioning (RP) [23]. This is a central C-SAR method in AB that proved very useful in the mechanistic analysis of complex BAs [32-34].

5.1. Recursive Partitioning (RP)

RP is a multi-step statistical procedure that can be viewed as an automated "multiple baseline" method. At each step of RP, all descriptors are sequentially analyzed in order to find the best criterion for splitting compounds into two classes ("active" and "inactive"). The best criterion is given by a structural or physicochemical descriptor with a certain cut-off value that produces the highest statistical significance. For example, one could obtain "log $P < 0.5$" or "No of COOH groups > 1". This procedure is repeated until no statistically significant splitting of new classes of compounds is possible. The result is a partitioning tree that needs to be mechanistically interpreted. One can easily obtain a large variety of "statistical" RP trees that group compounds into classes using many alternative criteria. Which criterion is correct when statistical significance is indifferent? The answer can only be obtained using a multi-step C-SAR analysis with sequential identification of the following effects: (i) physicochemical profiles of active and inactive compounds, (ii) "small structural" effects, such as chemical reactivity, (iii) "large structural effects" - the largest possible skeletons that are common to active compounds, (iv) construction of "pharmacophores" – structural constructs with site-specific generalizations. This order is strictly defined by the constructionist approach. The first three steps gradually reduce multiple alternative explanations, whereas the last step aims at increasing the generality of considerations.

5.1.1. First Step: Physicochemical Profiling (PP)

PP implies RP (or discriminant) analysis with various physicochemical properties. The result is a set of physicochemical cut-off values that can be compared to multiple Hansch equations. Similarly to the Hansch analysis, any PP is only useful if it satisfies the following conditions: (i) it employs reliable parameters, (ii) follows common knowledge about the underlying mechanisms, (iii) does not produce unexplained deviations. Any physicochemical explanations are much more general than sub-structural explanations, so PP must precede any sub-structural analysis. Fig. (**5-A**) shows an example of PP for human intestinal permeability [33].

Here T*PSA* and *MW* describe the entire structural space, facilitating a substructure-specific analysis of *HIA*. All compounds were subdivided into three types of permeability – paracellular, "non-restricted" and "restricted". The usefulness of this description can be seen from the following example. Klopman and co-workers [66] identified polar sugar skeletons in amikacin and neomycin as fragments that prohibit *HIA* ("biophobes"). The parent compounds – amikacin and neomycin - have very high T*PSA*, falling into the "non-permeable" region in Fig. (**5-A**). So the "biophobic" effect of respective sugar skeletons can be excluded from considerations at the earliest stages of *HIA* analysis. Fig. (**5-B**) shows an example of PP for P-gp substrate specificity [34]. Here *MW* and electrostatic charge (coming from the ionization of strong acids and bases) determine the regions for active and inactive compounds. The question marks designate the regions that require further substructure-specific analysis. Essentially, these are the only regions where 3D modeling may lead to any rational hypotheses. (Derivation of such hypotheses can be compared to structure-based design that is preceded by property-based design [25].) Both examples in Fig. (**5**) represent simplified versions of PP for the sake of simple visualization. In a general case scenario, any PP should be "multidimensional", representing permissible and non-permissible ranges for many different properties. Even if some properties may seem to be non-essential, a definition of their variability ranges

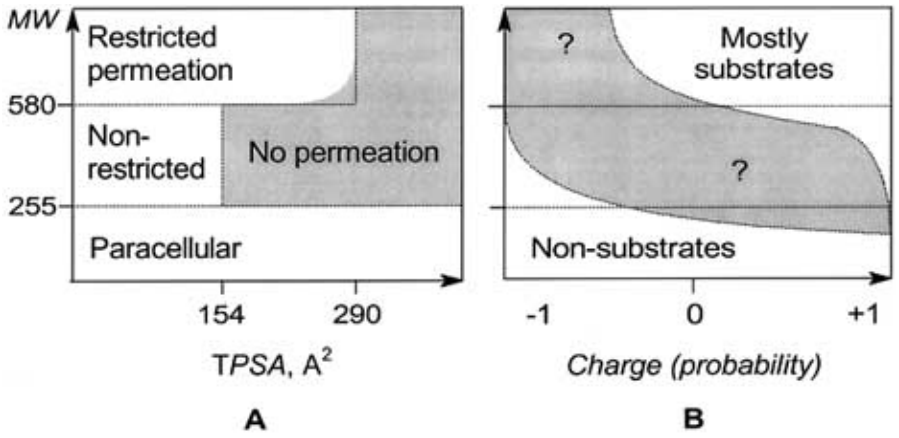

Fig. (5). Simplified PP for human intestinal permeability (A) and P-gp substrate specificity (B). In the latter case approximate *MW* cut-off values are shown.

may still prove meaningful, as it characterizes the diversity of the training set. Frequently multiple properties are combined into principal components using PCA, although this is not always permissible. For example, consider the analysis of *HIA* for highly polar compounds, assuming that T*PSA* and *MW* are inter-correlated. In this case PCA may combine these properties into one component, resulting in the loss of mechanistic reasoning.

5.1.2. Second Step: Identification of Small Structural Effects

These are reactive combinations of small functional groups that can be most easily identified using the IC-based and atom chain parameters shown in Fig. (**2-A,B**). Klopman used atom chains for identifying reactive skeletons in many types of BAs [20,21,66]. The most recent study involved receptors related to Parkinson's disease (dopamine receptor, MAO-B, N-methyl-D-Aspartate receptor) [67]. This analysis was not preceded by PP, as the analyzed activities were clearly defined by the sub-structural effects. A much more common situation is when such clarity is completely lacking. For example, consider the case of *HIA* [33]. Here PP was used to define non-specific permeability. Small structural effects were related to quaternary nitrogens (permanent charges that prohibit permeation) and biphosphonates (chelating ability towards Ca^{2+}). Some effects (e.g., active transport) could only be captured by a "manual" analysis of outliers. These results have clearly shown that good statistical significance of structural parameters does not automatically mean good mechanistic explanations.

5.1.3. Third Step: Identification of Largest Active Skeletons

These denote the largest possible skeletons that are typical of active compounds. Any diverse compounds can be viewed as a mixture of many smaller subsets representing "biologically similar" compounds. For example, natural compounds consist of analogs of peptides, alkaloids, etc. The same can be said of any therapeutic drugs that consist of multiple "narrow" chemical classes. Each class can be characterized by a certain skeleton that is much more meaningful than any statistically derived pharmacophore. If statistical results reveal no major difference between the "biological" and "non-biological" skeletons, the biological skeleton is always preferable. It is therefore important to identify such skeletons *prior* to the construction of "statistical" pharmacophores. P-gp substrate specificity is a good example. Seelig [68] identified several "pharmacophores" based on two- and three- point models. A verification of this idea revealed many alternative models with equally good statistical significance [34]. Many of these models could be eliminated by simple PP, as shown in Fig. (**5-B**). Among the remaining compounds, the highest biological meaning was obtained using common skeletons of the known P-gp substrates (analogs of peptides, taxanes, colchicines, anthracyclines, vinca alkaloids, etc.). Fig. (**6-A**) shows an example of the analogs of colchicines.

The identification of such large skeletons can be performed in two ways. The first is by means of the automated generation of large atom chains driven by the statistical significance of the smaller chains. Small "active" chains (that are typical of active compounds) are gradually enlarged, whereas chains that are not "active" are discarded. So we can identify skeletons of up to 15 atoms [29,32]. The second way is based on "rigid scaffolds" [69] that are shown in Fig. (**2-D**). Each ring skeleton can be used as a "large building block" in a dynamic construction of active skeletons [29]. Obviously, there are many situations when neither of these approaches will work. Fig. (**6-B**) shows an example of vinblastine and other similar natural compounds. Here the largest active skeleton proved to be very small compared to the actual molecules, clearly pointing to the current limitations of C-SAR analysis. Fig. (**6**) can also serve as an example of a much broader problem in C-SAR related to the critical assessment of the obtained results. The result (A) is meaningful, as it provides good explanations leading to high-quality predictions in a narrow structural space. The obtained skeleton can be further generalized in order to enhance its predictive power. The result (B) is not meaningful, as it provides no explanations. So it should be discarded from any further considerations.

5.1.4. Fourth Step – "Pharmacophore" Construction

This step implies a site-specific generalization of the active skeletons (small and large) that were identified in the previous steps. So we expand their predictive power, while

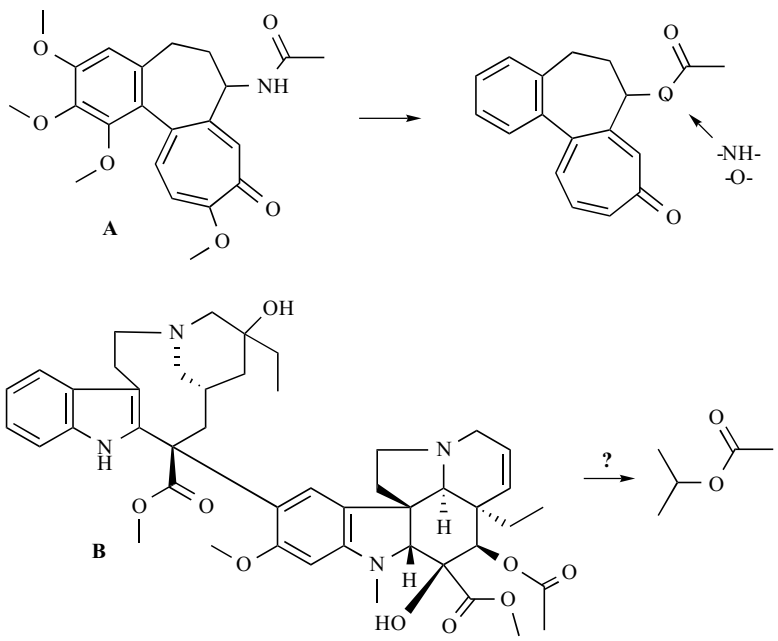

Fig. (6). Examples of identification of the "maximum active" skeletons using large atom chains and rigid scaffolds. A – analogs of colchicine, B – vinblastine and similar natural compounds.

preserving mechanistic reasoning. This means a "physicochemical induction" that can be compared to Hansch and Leo's generalization of CLOGP parameters. Obviously, this induction should only be applied to the skeletons that are mechanistically meaningful. In the case of small fragments, generalizations are achieved using simple generic atoms with variable topological properties (e.g., aromaticity, branching, etc.). For example, cholinesterase inhibiting groups can be represented as follows: (i) $R_3P{=}X$, where R – not OH, NH, SH, X - =O, =S, (ii) R-OCON<, where R – aromatic or oxime [32]. Similar substructures constitute the essence of the "knowledge bases" used in predictive toxicology [32,70]. In the case of large structural skeletons, generalizations must be started from the definition of the major "pharmacophoric domains" – ionization centers, electron donors, rigid rings and flexible chains. Each of these centers must be clearly defined, which is not always easy to do. For example, consider Fig. (**6-A**) - is the 7-member aliphatic ring essential for P-gp substrate specificity? If not, then the colchicine-type skeleton can be generalized as follows:

[aromatic ring] – [aliphatic chain of 2-3 atoms] – [e-donor].

This type of "pharmacophore" is very broad, so each domain must be further specified using various structural, topological and physicochemical definitions. For example, the replacement of an abstract "e-donor" with a small generic substructure (-XCOR, X is NH or O) can explain the activity of both structures in Fig. (**6**). This hypothesis can only be accepted if it produces high statistical significance – this is where statistical induction can be useful. This kind of inductive approach must be used for each "pharmacophoric domain". Can an aromatic ring involve any heteroatoms? Can an aliphatic chain involve bulky substituents? All hypotheses that are statistically meaningful have to be expressed in terms of either generic substructures or physicochemical cut-off values. Fig. (**7**) shows an example of the possible result.

Here the "2D pharmacophore" represents a combination of three structural segments with clearly defined structural variation limits. Each segment has to be characterized individually, using structural or physicochemical parameters that may not be applicable in the other segments. The situation is quite similar to PP, but it is now done at the site-specific level. The resulting 2D pharmacophore can be further converted into a 3D model in order to account for "delicate" conformational effects. Any 2D explanations are much more general than 3D explanations, so derivation of 2D pharmacophores must precede the derivation of 3D pharmacophores. If this is not done, the obtained 3D models resemble solutions of ill-conditioned systems of equations that lead to chance correlations. To the best of our knowledge, so far site-specific PP was not explicitly used in any 3D studies. De Groot and Ekins [72,73] provided an example of the implicit use of this approach in predicting P450 mediated metabolism. Each of the analyzed compounds received careful site-specific consideration, costing many years of "non-statistical" analysis of CYP 2D6 specificity. This can be viewed as site-specific PP that was facilitated by *a priori* mechanistic knowledge. The subsequent 3D modeling can be viewed as a fine-tuning of various conformational effects.

6. CONCLUSIONS

Using chemical and biological information as simple parameters that facilitate the gradual reduction of multiple explanations constitutes the essence in the mechanistic approach. Any type of chemical or biological information can be expressed as certain "mechanistic" descriptors. The logical meaning of these descriptors must correspond to the complexity of the analyzed properties. If we deal with "simple" log *P*, an analysis in terms of intra-molecular interactions is sufficient. But if we deal with complex *in vivo* data, any known ADME/Tox effects should be

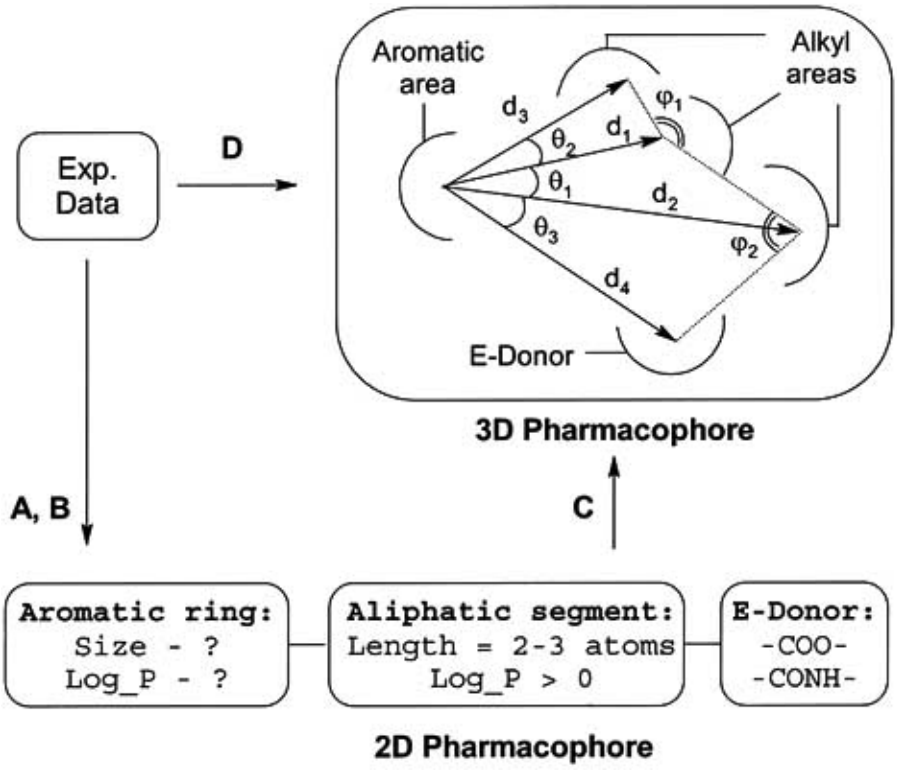

Fig. (7). Comparison of 2D and 3D pharmacophore models for vinblastine. The 2D model is only an illustration of possibilities that may not correspond to reality. The 3D model was taken from literature [71]. Steps A – D correspond to those in Fig. (**1**).

represented as "indicator" parameters [32]. By converting our knowledge into simple parameters we identify compounds that deviate from the existing knowledge. An analysis of these deviations is the most valuable source of new mechanistic information. This information can only be obtained in a multi-step deductive approach, as the number of alternative explanations is always greater than (or comparable to) the number of experimental observations. Analysis must start from the simplest compounds, with a gradual increase in structural complexity. Explanations must start from the simplest considerations, with a gradual increase in structural specificity:

Physicochemical > 2D structural > 3D (conformational) > (Biochemical)

When the available information reaches a certain critical threshold, any 2D (and 3D) structural explanations can be generalized using physicochemical, chemical or even biological considerations. This leads to the increased predictive power of our models, and represents the major source of good statistical significance in our correlations. So far this scenario has not been very popular, as good statistical significance was usually obtained using various single-step (statistical and similarity) approaches. This indicates a potentially big problem in cheminformatics, as the blind optimization of statistical significance is the "dead-end" of any mechanistic analysis. More than two decades ago Hansch [2] pointed out that "If QSAR accomplishes nothing more than to get chemists to take a more thoughtful and logical attitude in derivatizing a lead compound, it will have made an important contribution". In the case of diverse compounds, any single-step statistical optimization has a reverse effect, as it discourages us from thinking critically about the underlying biological mechanisms.

7. ACKNOWLEDGEMENTS

The authors are thankful to Dr. Derek Reynolds (Reytek Limited, UK) for inspiring discussions; Dr. Giulia Caron for valuable comments; and Darius Ross for his careful proofreading.

ABREVIATIONS

AB	=	Algorithm builder
ANNs	=	Artificial neural networks
BA	=	Biological activity
C-SAR	=	Classification structure-activity relationship
FM	=	Fragmental method
HIA	=	Human intestinal absorption
LFER	=	Linear free energy relationship
Log *D*	=	Octanol/water distribution coefficient for electrolytes at a certain pH value
Log *P*	=	Octanol/water partition coefficient for the neutral species
Log S_w	=	Logarithm of characteristic solubility
MW	=	Molar weight

MV = Molar volume

MR = Molar refraction

mp = Melting point

PCA = Principal component analysis

PLS = Partial least squares

pK_a = Acid-base ionization constant

PP = Physicochemical profiling

RP = Recursive Partitioning

QSAR = Quantitative structure-activity relationship

QSPR = Quantitative structure-property relationship

T*PSA* = Topological polar surface area.

REFERENCES

[1] Albert, A. Ed.; *Selective Toxicity with Special Reference to Chemotherapy*. Methuen & Co.: London, **1951**.

[2] Hansch, C. In *Biological Activity and Chemical Structure*; Kverling Buisman, J.A.; Ed.; Elsevier: Amsterdam, **1977**; pp 47-61.

[3] Kubinyi, H. *Prog. Drug. Res.*, **1979**, *23*, 97-198.

[4] Kubinyi, H. In: Wolff, M. E. (Ed.), *Burger´s Medicinal Chemstry and Drug Discovery*, 5th Edition, Vol. I, Principles and Practice, John Wiley & Sons, New York, **1995**, pp. 497-571.

[5] Kubinyi, H. In: Schleyer, P. V. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer III, H. F., Schreiner, P. R. (Eds.), *The Encyclopedia of Computational Chemistry*, Vol. 4, John Wiley & Sons, Chichester, **1998**, pp. 2309-2320.

[6] Kubinyi, H. Ed.; 3D QSAR in drug design. Theory, methods and applications, ESCOM Science Publishers, **1993**.

[7] Kubinyi, H. *Drug Discovery Today*, **1997**, *2*, 538-546.

[8] Kubinyi, H. In: Schleyer, P. V. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer III, H. F., and Schreiner, P. R. (Eds.), *The Encyclopedia of Computational Chemistry*, Vol. 4, John Wiley & Sons, Chichester, **1998**, pp. 449-460.

[9] Kubinyi, H. *Current Drug Discovery*, **2001**, October, 9-11.

[10] Bakken, G.A.; Jurs, P.C. *J. Med. Chem.*, **2000**, *43*, 4534-4541.

[11] Rabow, A.A.; Shoemaker, R.H.; Sausville, E.A.; Covell, D.G. *J. Med. Chem.*, **2002**, *45*, 818-840.

[12] Pepperrell, C. *Three-Dimensional Chemical Similarity Searching*, Wiley, Chichester, **1994**.

[13] Downs, G. M.; Willett, P. M. *Rev. Comput. Chem.,* **1995**, *7*, 1-66.

[14] Willett, P. *J. Chem. Inf. Comp. Sci.,* **1998**, *38*, 983-996.

[15] Andrews, C.W.; Bennett, L.; Yu, L.X. *Pharm. Res.,* **2000**, *17*, 639-644.

[16] Raevsky, O.; Trepalin, S.V.; Trepalina, H.P.; Gerasimenko, V.A. *J. Chem. Inf. Comput. Sci.,* **2002**, *42*, 540-549.

[17] Ghuloum, A.M.; Sage, C.R.; Jain, A.N. *J. Med. Chem.,* **1999**, *42*, 1739-1748.

[18] Anzali, S.; Barnickel, G.; Cezanne, B.; Krug, M.; Filimonov, D.; Poroikov, V. *J. Med. Chem.,* **2001**, *44*, 2432-2437.

[19] HQSAR (computer program) available from Tripos, Inc., 1699 South Hanley Road, St. Louis, MO 63144-2913, USA. www.tripos.com/sciTech/inSilicoDisc/media/LITCTR/HQSAR_AP.PDF.

[20] Klopman, G. *Quant. Struct. Act. Relat.,* **1992**, *11*, 176-184.

[21] Rosenkranz, H.S.; Cunningham A.R.; Zhang, Y.P.; Klopman, G. *SAR QSAR Environ. Res.*, **1999**, *10*, 263-276.

[22] Judson, P. N. *J. Chem. Inf. Comp. Sci.,* **1994**, *34*, 148-153.

[23] Rusinko III, A.; Farmen, M.W.; Lambert, C.G.; Brown, P.L.; Young, S.S. *J. Chem. Inf. Comp. Sci.,* **1999**, *39*. 1017-1026.

[24] Izrailev, S.; Agrafiotis, D. *J. Chem. Inf. Comput. Sci.,* **2001**, *41*, 176-180.

[25] van de Waterbeemd, H.; Smith, D.A.; Beaumont, K.; Walker, D.K. *J. Med. Chem.,* **2001**, *44*, 1313-1333 .

[26] Testa, B., Crivori, P., Reist, M., Carrupt, P.-A. *Persp. Drug Disc. Design,* **2000**, *19*, 179-211.

[27] Hansch, C.; Kurup, A.; Garg, R.; Gao, H. *Chem. Rev.,* **2001**, *101*, 619-672.

[28] Japertas, P.; Didziapetris, R.; Petrauskas, A. *Quant. Struct.-Act. Relat*., **2002,** *21*, 23-37.

[29] Algorithm Builder (computer program) available from Pharma Algorithms, Inc., Tauro 12, Vilnius, 2001 Lithuania; www.ap-algorithms.com.

[30] AB/Ionization (computer program) available from Pharma Algorithms, Inc., Tauro 12, Vilnius, 2001 Lithuania; www.ap-algorithms.com .

[31] AB/Solubility (computer program) available from Pharma Algorithms, Inc., Tauro 12, Vilnius, 2001 Lithuania; www.ap-algorithms.com .

[32] Zmuidinavicius, D.; Japertas, P.; Petrauskas, A.; Didziapetris, R. *Curr. Top. Med. Chem.*, **2003**, *3*, 1301-1314.

[33] Zmuidinavicius, D.; Didziapetris, R.; Japertas, P.; Avdeef, A.; Petrauskas, A. *J. Pharm. Sci.*, **2003**, *92*, 621-633.

[34] Didziapetris, R.; Japertas, P.; Petrauskas, A. In *Drug Transporters. Can They Be Modelled?* The 14th European Symposium on Quantitative Structure-Activity Relationships, Workshop I, Bournemouth, UK; **2002**.

[35] Lipnick, R. L. *Sci. Totat. Environ.,* **1991**, *109-110*, 131-153.

[36] Hansch, C.; Leo, A. Eds.; Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley: New York, **1979**.

[37] Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR: Hydrophobic, Electronic and Steric Constants*, ACS: Washington, DC, **1995**.

[38] Leo, A. *Chem. Rev.*, **1993**, *93*, 1281-1306.

[39] Tetko, I.V.; Tanchuk, V.Y.; Villa, A.E. *J. Chem. Inf. Comp. Sci.*, **2001**, *41*, 1407-1421.

[40] Howard, P.; Meylan, W.M. In Quantitative Structure-Activity Relationships in Environmental Sciences, Chen, F.; Schuurman, G.; Eds.; SETAC Special Publication Series, **1977**, pp. 185-205.

[41] Mannhold, R.; Dross, K. *Quant. Struct.-Act. Relat.,* **1996**, *15*, 403-409.

[42] Mannhold, R.; Petrauskas, A. *Quant. Struct. Act. Relat. Comb. Sci.*, **2003**, *22*, 466-475.

[43] Rekker, R.F. *The Hydrophobic Fragmental Constant, Pharmacochemistry Library,* Elsevier: Amsterdam, Vol. 1; **1977**.

[44] Rekker, R.F.; Mannhold, R. *Calculation of Drug Lipophilicity*, VCH: Weinheim, **1992**.

[45] Meylan, W.M.; Howard, P.H. *Persp. In Drug Discovery and Design,* **2000**, *19*, 67-84.

[46] Avdeev, A. *Curr. Top. Med. Chem.*, **2001**, *1*, 277-352.

[47] Kubinyi, H. In *Pharmacokinetic Optimization in drug research*, Testa, B.; van de Waterbeemd, H.; Folkers, G.; Guy, R., Eds., Wiley, **2001**, pp. 513-524.

[48] Lombardo, F.; Obach, R.S.; Shalaeva, M.Y.; Gao, F. *J. Med. Chem.*, **2002**, *45*, 2867-2876.

[49] da Silva, C.O.; da Silva, E.C.; Nascimento, M.A.C. *J. Phys. Chem. A*, **1999**, *103*, 11194-11199.

[50] Citra, M.J. *Chemosphere*, **1999**, *38*, 191-206.

[51] Klopman, G.; Fercu, D. *J. Comput. Chem.*, **1994**, *15*, 1041-1050.

[52] Xing, L.; Glen, R.C. *J. Chem. Inf. Comput. Sci.*, **2002**, *42*, 796-805.

[53] Bremser, W. *Anal. Chim. Acta*, **1978**, *103*, 355-365.

[54] Perrin, D.D.; Dempsey, B.; Serjeant, E.P., pK_a *Prediction for Organic Acids and Bases*, Chapman and Hall: London and New York, **1981**.

[55] Tsantili-Kakoulidou, A.; Panderi, I.; Csizmadia, F.; Darvas, F. *J. Phram. Sci.*, **1997**, *86*, 1173-1179.

[56] Hilal, S.H.; Carreira, L.A.; Karickhoff, S.W., In *Quantitative Treatments of Solute/Solvent Interactions: Theoretical and Computational Chemistry*, Elsevier: New York, **1994**; Vol. 1, pp. 291-353.

[57] ACD/pK_a (computer program) available from Advanced Chemistry Development, Inc., 90 Adelaide Street West, Suite 702, Toronto, Ontario M5H 3V9, Canada; www.acdlabs.com.

[58] ACD/Solubility (computer program) available from Advanced Chemistry Development, Inc., 90 Adelaide Street West, Suite 702, Toronto, Ontario M5H 3V9, Canada; www.acdlabs.com.

[59] Klopman, G.; Zhu, H. *J. Chem. Inf. Comput. Sci.,* **2001**, *41*, 439-445.

[60] Yalkowsky, S.H.; Banerjee, S. Aqueous solubility methods of estimation for organic compounds. Marcel Dekker: New York, **1992**.

[61] Ran, Y.; Yalkowsky, S.H. *J. Chem. Inf. Comput. Sci.,* **2001**, *41*, 354-357.
[62] Abraham, M.H.; Lee, J. *J. Pharm. Sci.*, **1999**, *88*, 868-880.
[63] Svozil, D.; Sevcik, J.G.K.; Kvasnicka, K. *J. Chem. Inf. Comp. Sci.*, **1997**, *37*, 338-342.
[64] Hansch, C; Quinlan, J.E.; Lawrence, G.L. *J. Org. Chem.* **1968**, *33*, 347-350.
[65] Lipnick, R. L. In: *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W.; Devillers, J.; Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, **1990**; pp 129-144.
[66] Klopman, G.; Stefan, L.R.; Saiakhov, R.D. *Eur. J. Pharm. Sci.*, **2002**, *17*, 253-263.
[67] Klopman, G.; Sedykh, A. *BMC Pharmacology*, **2002**, *2*, 1-11.
[68] Seelig, A. *Eur. J. Biochem.*, **1998**, *251*, 252-261.
[69] Xu, J.; Stevenson, J. *J. Chem. Inf. Comput. Sci*., **2000**, *40*, 1177-1187.
[70] Greene, N.; Judson, P.N.; Langowski, J.J.; Marchant, C.A. *SAR QSAR Environ. Res*. **1999**, *10*, 299-314.
[71] Garrigues, A.; Loiseau, N.; Delaforge, M.; Ferte, J.; Garrigos, M.; Andre, F.; Orlowski, S. *Mol. Pharmacol.,* **2002**, *62*, 1288-1298.
[72] de Groot, M.J.; Ekins, S. *Adv. Drug Deliv. Rev.*, **2002**, *54*, 367-83.
[73] Ekins, S; de Groot, M.J.; Jones, J.P. *Drug Metab. Dispos.*, **2001**, *7*, 936-944.