

QSAR: dead or alive?

Arthur M. Doweyko

Received: 27 November 2007 / Accepted: 10 December 2007 / Published online: 9 January 2008
© Springer Science+Business Media B.V. 2007

Abstract This perspective concerns the methods employed within the current drug discovery community to develop predictive quantitative structure–activity relationships (QSAR). Specifically, a number of cautions are provided which may circumvent misuse and misunderstanding of the technique. Ignorance of such caveats has led to a discouraging tendency of the methods to result in poorly predictive models. Among these pitfalls are the fondness with which we associate correlation with causation, the mesmerizing influence of large numbers of molecular descriptors, the incessant misuse of the leave-one-out paradigm, and finally, the QSAR enigma wherein model predictivity is not a necessary component of a model's usefulness.

Keywords Correlation · Descriptors · Drug design · QSAR · Quantitative structure–activity relationships

Introduction

The concept of quantitative structure–activity relationships (QSAR) is inherently associated with optimism, a mindset ever hopeful for predictive correlations and the prospects of novel insight or hypothesis. However, lately the concept engenders quite the opposite reaction from the scientific community-at-large, a negative view which is not entirely without merit. In contrast to the pioneering work of Hansch and Fujita [1], the modern QSAR era has witnessed a vast

number of studies (2D and 3D) with sufficiently poor predictive qualities to underscore a growing shadow of doubt on an ever-darkening correlative landscape. Is this an actual failing of QSAR? Or, is there something else afoot? This perspective will make an attempt at identifying those segments of QSAR methodology that may have lent themselves to misuse, misunderstanding and, of course, mistakes.

Correlation and causation

There are numerous examples of observations which absolutely involve no causality but are nonetheless significantly correlated. For example, Sies [2], presumably with tongue in cheek, wrote to the editors of *Nature*: “Sir—There is concern in West Germany over the falling birth rate. The accompanying graph might suggest a solution that every child knows makes sense.” That graph is illustrated in Fig. 1 along with a correlation plot. There is indeed a strong $r^2 = 0.99$ correlation between the numbers of brooding storks and newborn babies. In a child's world this certainly makes a lot of sense. However, in an adult's world (at least for most adults), causality of this sort might be somewhat suspect. A second example is one which shows a correlation between the US population and the number of civil executions (Fig. 2), [3, 4] suggesting that a decrease in civil execution activity is associated with higher population counts ($r^2 = 0.99$), with the putative hypothesis that the population increase is a consequence of the decreased number of executions! This almost sounds logical, until the magnitude of the numbers is considered. We react to such correlations with amusement, and are apt to quickly relegate them to coincidence and nothing more ... but what is it that permits us to do this? If the ‘brooding stork’ descriptor were some topological index

A. M. Doweyko (✉)
Bristol-Myers Squibb, Research and Development,
CADD Group, P.O. Box 4000, Princeton, NJ 08543, USA
e-mail: arthur.doweyko@bms.com

Fig. 1 A plot of the numbers of pairs of brooding storks and newborn babies in West Germany from 1960 to 1985. Representation of the data as a correlation plot (Inset)

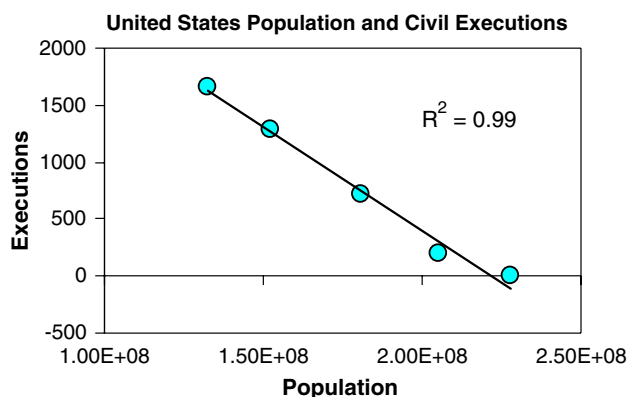
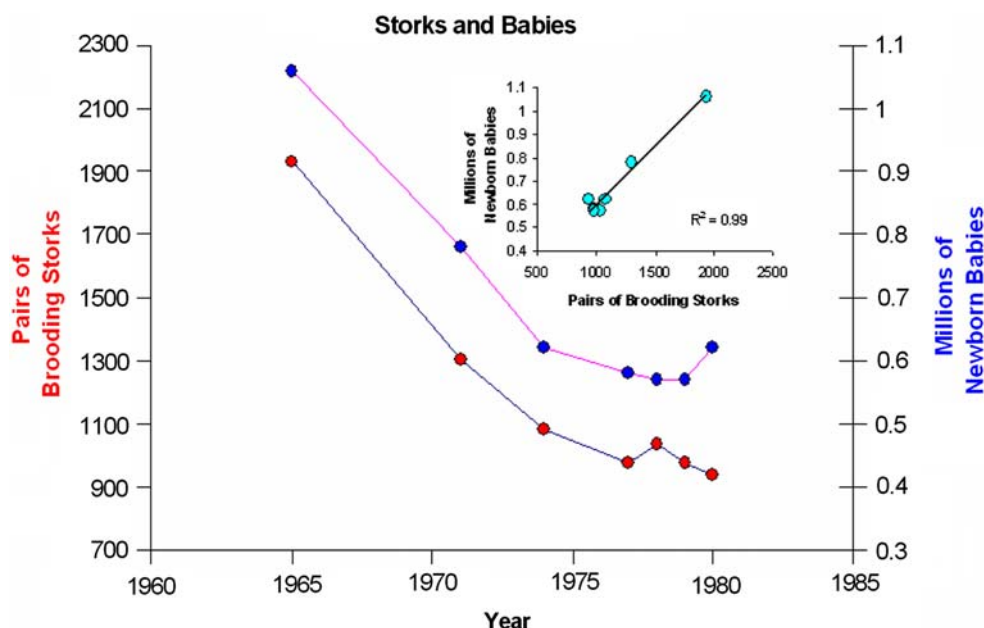


Fig. 2 United States population and civil executions (1940–1980, taken from Ref. [3])

and newborn babies were binding affinities would it be so easy to decide on causality? Our ability to distinguish between cause and coincidence in these cases is based on experience. In other words, we conduct a mental experiment on the two observations and decide on a likely outcome.

Sometimes more care needs to be taken when making observations. For example, back in 1897, Karl Pearson

discovered an interesting correlation when examining a large collection of skulls gathered from the Paris Catacombs [5, 6]. The goal of the assessment was to uncover a possible relationship between skull length and breadth. Specifically, if skull shape were constant, then such a comparison would yield a positive correlation. On the other hand, if skull volume were constant, then the correlation would be negative. So Karl went into the investigation with some pre-conceived hypotheses. The results are typified by the plot in Fig. 3. A reasonable positive correlation was found and that was consistent with the constant shape paradigm. However, after closer inspection Karl noticed that if the female skulls and male skulls were divided into separate categories, the correlation with shape disappeared. This remarkable real life case is a superb example of correlation analysis nearly gone astray. The same issue can easily be encountered in the extraction of possible correlations between activity and structure when examining diverse sets of compounds, or looking for trends in large databases while ignoring the effect of chemotype populations within such sets.

Experience, although quite useful at times, also lends itself to bias. A case in point involves sunspots, and the

Fig. 3 Left panel: An illustration of the relationship between skull length and width, exhibiting an apparent correlation. Right panel: Data segregated by sex (female shown in magenta, male shown in green) showing no significant correlation

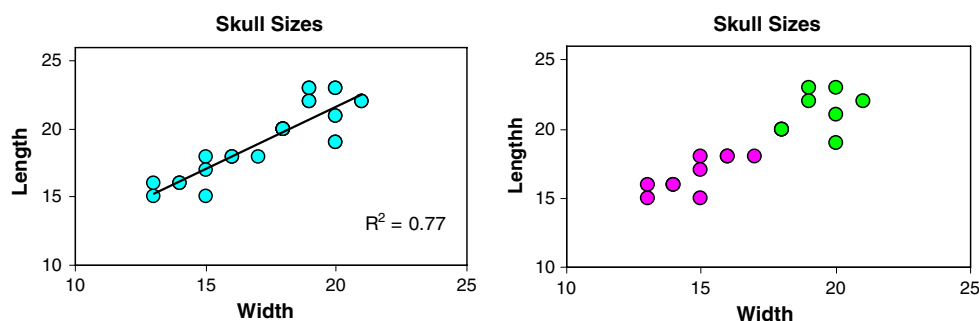
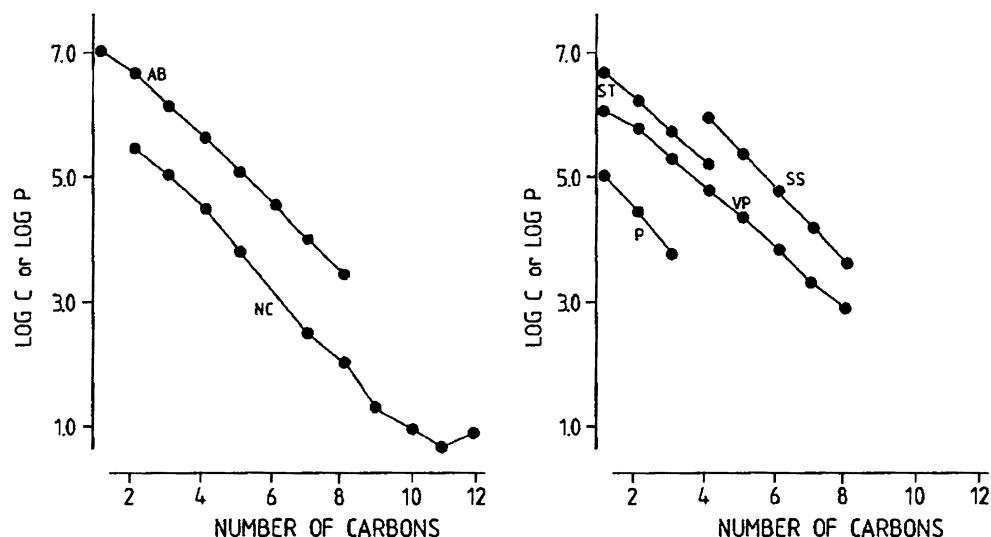


Fig. 4 Effects of *n*-alcohol chain length on some chemical and biological properties: (Left Panel) AB = log of concentration inhibiting *S. Typhosus*, NC = log of concentration causing narcosis in tadpoles. (Right Panel) ST = log of concentration lowering surface tension by a standard amount, SS = log of concentration producing a saturated solution, VP = log of vapor pressure, and P = log of partition coefficient between water and croton oil (data taken from Ref. [8])



bias is simply that sunspots are likely to be of little consequence other than causing some spectacular lights in the sky along with poor radio reception. Solar activity has been blamed for a number of events here on Earth including the epidemics of diphtheria and smallpox, weather patterns, revolutions, financial crises, road accidents and mining disasters. Interestingly, these conclusions were largely arrived at by conducting careful correlative analyses. There is a human tendency to attach causation to these phenomena, mainly due to impressive correlation statistics; it would appear that some of us lightly step over the basic tenets of the Scientific Method and conveniently forget about experimentally validating a hypothesis. An example of one such carefully conducted investigation attempts to address reports of heightened solar activity associated with cardiovascular events such as coagulation disorders, myocardial infarctions, stroke, arrhythmias and death [7]. C-Reactive protein (CRP) represents a major inflammation and acute phase marker in the progression of cardiovascular conditions. The results from over 25,000 serum CRP tests spanning a three year period were examined, leading to a striking correlation between CRP levels and solar geomagnetic data (GMA) as well as cosmic ray data (CRA). Regardless of bias to the contrary, these analyses do suggest a possible connection between our ionosphere and our health, and basically beg for an experimental follow-up to help define the mechanism behind such a connection.

Before we leave the subject of correlation-inferred causation and consider QSAR practices in drug design, there are a couple of examples of chemical structure correlations which may serve to underscore the wobbly nature of inference in molecular design. The trends depicted in Fig. 4 illustrate the effect of alkyl alcohol chain length on some chemical and biological properties [8]. For example, the growth inhibition of *S. Typhosus* (AB) and narcosis in

tadpoles (NC) is increased as the number of carbons increases. One might conclude that this is a structural effect. However, other decreasing trends in the series are evident: vapor pressure (VP), the partition coefficient between water and cottonseed oil (P), surface tension (ST) and solubility (concentration needed to reach saturation, SS). It is possible that one or more of these effects are causative, and it is further possible that an as yet unmeasured property of the series may be involved. In fact, there are many calculated properties of these alcohols that could be correlated to the observed biological effects. Until an experiment is designed to test a hypothesis, we will never know the mechanistic foundation, and therefore, will also not be assured of a predictive correlation. A contrasting example of cause and effect is shown in Fig. 5, wherein the activity of a series of sulfanilamides is plotted against measured pKa [9]. The structure–activity relationship is represented by a parabolic curve with

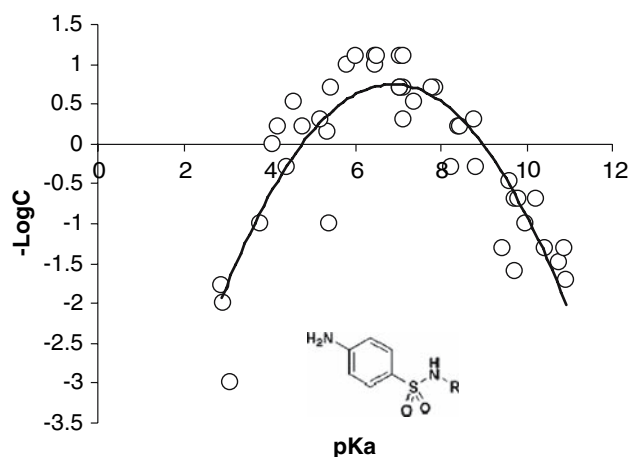
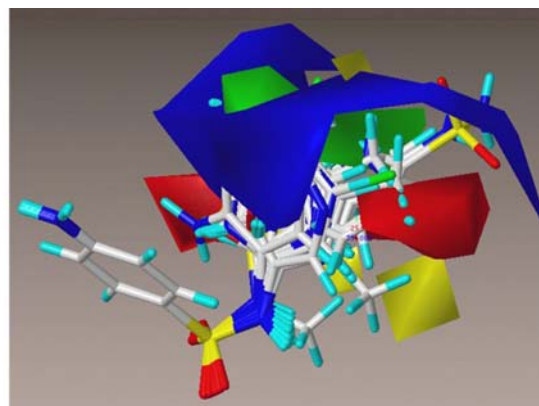
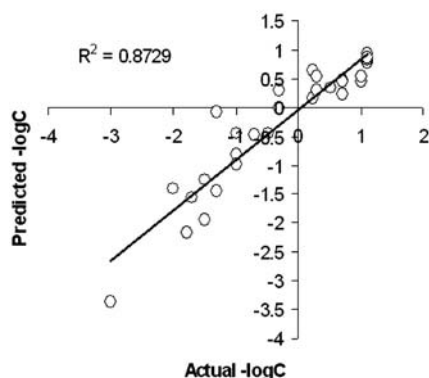


Fig. 5 The activity of sulfanilamides, expressed as the logarithm of the reciprocal of the concentration inhibiting the growth of *E. coli* plotted against measured pKa (taken from Refs. [8, 9])

Fig. 6 *Left:* Predicted (y-axis) versus Actual (x-axis) activities of the sulfanilamides generated with a standard CoMFA analysis using steric and electrostatic fields. *Right:* CoMFA steric (green and yellow) and electrostatic (blue and red) fields shown as contours mapped unto overlaid sulfanilamides

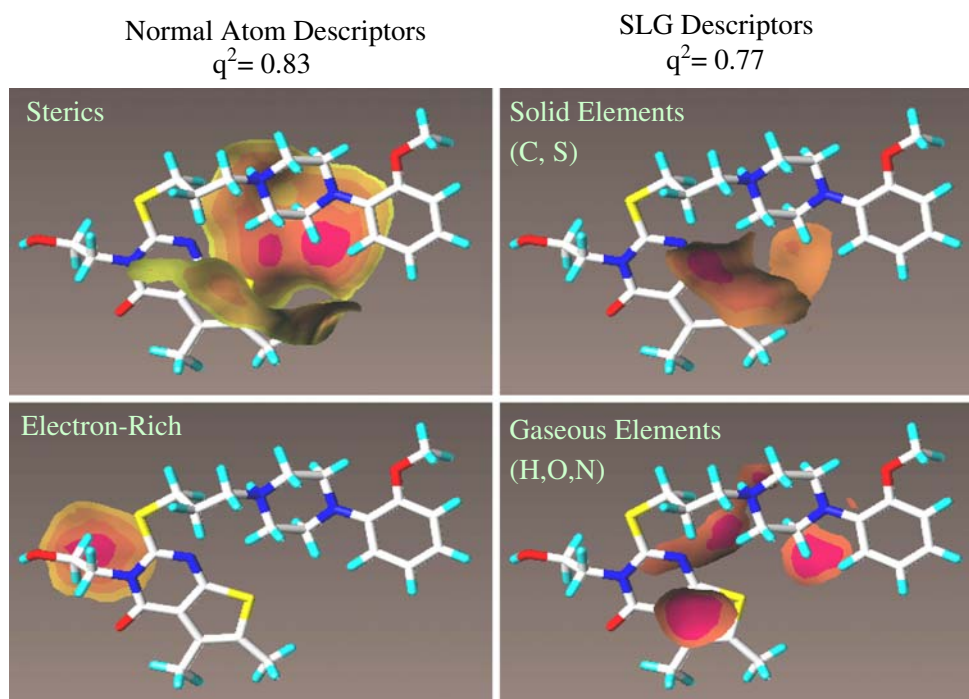


optimum bacteriostatic activity at pKa values between 6 and 7. The authors suggest that the mechanism at play entails bacterial cell wall permeability, and that only those sulfanilamides with just the right ratio of ionized to un-ionized forms can penetrate into the cell. This correlation suggests that there may not be an actual structure–activity relationship, since regardless of the nature of the *N*-substitution on these sulfanilamides (50 *N*-substituted analogues tested), their activities were dependent on pKa and not the structure of their *N*-substituents. Having said that, I took the opportunity to submit these data to a 3D-QSAR analysis, using CoMFA, and obtained the results depicted in Fig. 6. Depending on the number of PLS components used, the training set afforded r^2 values of 0.6–0.8. The model suggested a large sterically favorable (green) region adjacent to a region favorable towards positive electrostatic (blue) components of the molecule. These interaction fields explain

the variation in the observed bacteriostatic activity without the need to consider pKa. Thus, we have two fundamentally different potential explanations for the biological variability of these sulfanilamides, both statistically valid, with no way to tease them apart except by experiment.

These days we can find ourselves entangled in a descriptor jungle, unsure of how many and what types to use. These descriptors are often correlationally ensnared with one another, making it even more difficult to identify which sets relate to causality. A case in point is illustrated for 14 5HT1A ligands [10]. These data were subjected to a 3D-QSAR HASL analysis wherein atoms could be arbitrarily defined. When they were defined in a classical manner (three types: electron-rich, electron-poor and electron-neutral), the resulting q^2 (LOO) was 0.83 and the regions identified as important for activity coincided with a reasonable interpretation of the known SAR (Fig. 7) [11].

Fig. 7 HASL models of the 5HT1A data set. Normal atom descriptor model shown on left, SLG (solid–liquid–gas) descriptor model shown on right. Highlighted contours contribute positively to the predicted activity



However, if some other classification paradigm was adopted, for example, the capricious use of the elemental states (solid, liquid or gas) at room temperature as atom-descriptors, it was still possible to generate a mathematically plausible 3D-QSAR model ($q^2 = 0.77$). Here we have a model which points to the importance of atoms that are naturally gaseous or solid occupying specific molecular locations! These regions of importance to the model defy mechanistic interpretation, and illustrate the remarkably misleading results one can obtain when using uninterpretable descriptors.

Early days

Efforts to map molecular property to biological activity started about 140 years ago [12].

In his 1863 thesis work at the University of Strasbourg A. F. A. Cros noted a relationship between the water solubility of alcohols and their toxicity to mammals [13]. A few years later, Crum-Brown and Fraser [14] suggested the following general equation:

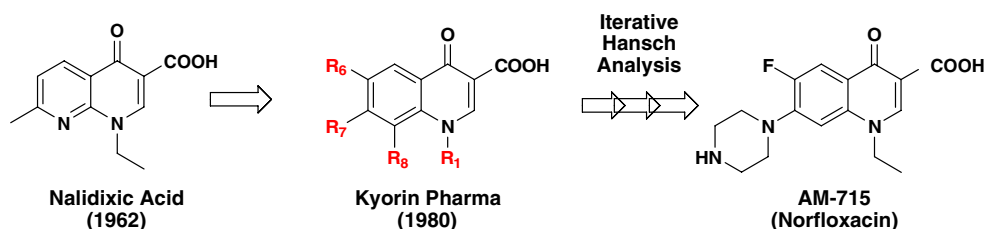
$$\Phi = f(\text{Constitution}) \quad (1)$$

Essentially, it was hypothesized that the structure of a molecule had something to do with its effect on biological systems. Despite this ground-breaking thesis, it was not until a quarter of a century later that Richet was able to formally demonstrate a correlation between structure and activity, wherein the ever-popular endpoint of toxicity of a series of simple organic compounds was found inversely related to solubility in water [15]. The partitioning behavior of molecules between organic solvent and water fascinated scientists, as this measurable ratio was increasingly found to be very significant in explaining (correlating with) the activities of many different types of organic molecules

and their observed biological activities. Once Hammett introduced a mechanistic means to capture electronic effects (Hammett σ constant) [16], the stage was set for the next quantum leap championed by Free and Wilson [17] and Hansch and Fujita [1]. By using a combination of partition coefficients, Hammett-derived constants and indicator variables, it was possible to generate meaningful and predictive structure–activity relationships! The really fascinating thing here is that successful QSAR entailed the use of a limited number of descriptors many of which were measured properties. At the same time methods were already in development to more accurately estimate the partition coefficients without actually measuring them. The turbulent horizon of change was clearly visible as larger drug molecule databases begged for correlative analyses, more complicated molecules demanded more detailed structural descriptors, and faster computational methods ran on faster machines to metamorphose QSAR to the unnecessarily confused state in which it exists today. As a rough estimate, until the early 1980's, QSAR was both a correlative and a predictive tool. For example, no less than 40 different successful QSAR-guided syntheses were cited in a 1981 perspective wherein predictions led to the synthesis of novel and active compounds [18]. These examples included a wide variety of biological endpoints such as enzyme inhibitors, insecticides, antibacterials, CNS and oncology agents.

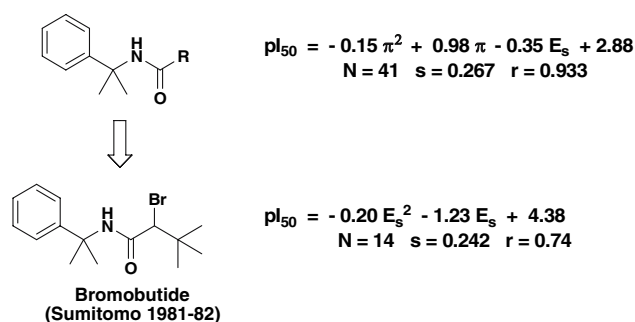
Norfloxacin, a synthetic, broad-spectrum antibacterial agent, represents an example of an actual marketed product discovered through QSAR analyses [19]. Scheme 1 illustrates the conceptual design flow followed by investigators at Kyorin Pharmaceuticals. Starting with nalidixic acid (1962) efforts were undertaken to determine the best combination of substituents about the quinoline ring system. Adopting a Hansch approach, AM-715 (Norfloxacin, 1980) was identified after selecting the best congeners with

Scheme 1 The development of the novel quinoline carboxylic acid anti-bacterial drug, AM-715 (Norfloxacin)



$$\begin{aligned} \text{Log}(1/\text{Mic}) = & -0.36 L_1^2 + 3.04 L_1 \\ & -2.50 E_s(6)^2 - 3.35 E(6) \\ & -1.02 B_4(8)^2 + 3.72 B_4(8) \\ & -0.21 \Sigma\pi(6,7,8)^2 - 0.48 \Sigma\pi(6,7,8) \\ & + 0.99 I_{(7)} - 0.73 I_{(7N-CO)} \\ & -0.68 \Sigma F(6,7,8) - 4.57 \end{aligned}$$

$$N = 71 \quad s = 0.274 \quad r = 0.964$$



Scheme 2 The development of the herbicide, S-47 (Bromobutide)

an optimal spectrum of activity, toxicity and cost of synthesis. The equation evolved from this QSAR analysis was based on 71 analogues and highlighted important structural and electronic features embodied in the Verloop STERIMOL descriptors, electronic, partitioning and indicator variables. A second example of a successful QSAR outcome is the herbicide, S-47 or Bromobutide, (Sumitomo Chemicals, 1981). Scheme 2 illustrates the structure–activity relationships developed during the course of its development, which also embody partition and electronic parameters. In both examples, QSAR approaches were used to guide the synthesis of one related structural series to the next.

Lately

Aside from the dramatic improvements in CPU speed and algorithm development, the greatest technological impact on modern QSAR has been the unbridled generation of molecular descriptors. This plethora of descriptors is both a wonder and a bane. As noted, we have progressed from descriptors that were simple to measure and understand (which generated QSAR equations with reasonable predictivity) to much more complicated ones designed to capture every nuance of molecular architecture and potential intermolecular interactions (which generate QSAR equations of questionable predictivity). The preference for fast descriptor calculation over measured physical attributes reflects a necessary requirement in drug discovery today, as large numbers of compounds are synthesized and tested in high-throughput mode. The implied correlation between poor QSAR performance and the staggering leap in descriptor calculation is not accidental. Binding affinity models can now be generated using a host of approaches (2D and 3D), each providing different sets of parameters which can appear to be important. Thus, a large number of statistically valid correlations are likely to leave the investigator at a loss to choose which, if any, equally valid, putative causative relationships exist. As if that was

not enough of a problem, the process of choosing a subset of correlated descriptors from a large pool is beset with the ‘Chance Factor’ effect so aptly described by Topliss and Edwards [20]. The inspiration behind their work stemmed from the observation that many researchers tended to provide statistics based on the final QSAR equation. Such criteria do not take into account how many independent variables were actually screened for possible inclusion in the equation. Clearly, the larger the number of possible independent variables is considered, the greater chance that an accidental correlation will occur, which is not at all reflected in the standard statistical criteria for the final equation. In an effort to identify the magnitude of such effects, a series of simulated QSAR studies employing random numbers was conducted. The risk of chance correlations was determined through a wide range of combinations of observations and screened variables using multiple-regression analyses. The trends shown in Fig. 8 summarize the dramatic effect that the number of independent variables has on the mean r^2 completely due to chance. For example, starting with 10 independent variables and 15 observations, the average number of variables entered by step-up multiple regression was 1.92 (not shown) with a mean r^2 of 0.46. If 20 independent variables

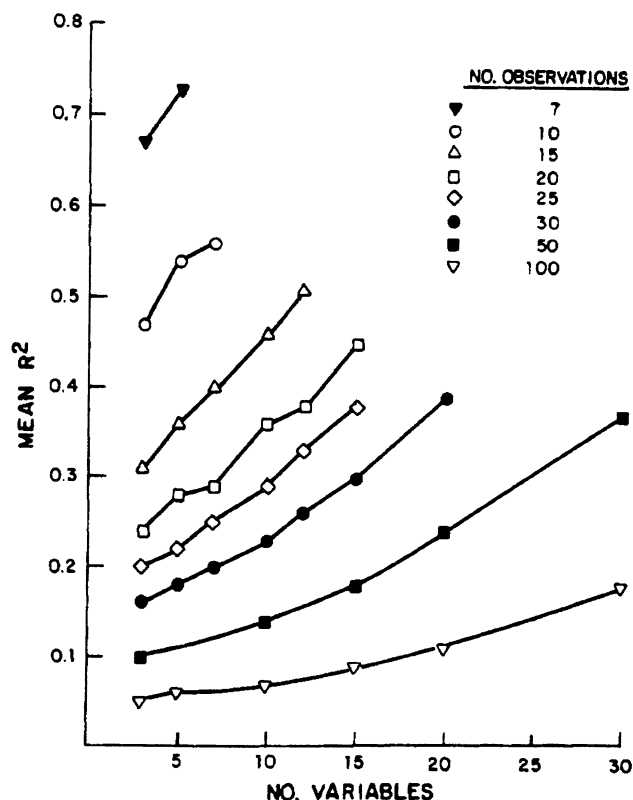


Fig. 8 Relationship between mean r^2 values and the number of variables screened using random number sets (figure reprinted from Ref. [20] by permission of the American Chemical Society)

are used, the mean r^2 value rises to 0.73! Of course, increasing the number of observations naturally decreases the degree of chance correlation. These results are extremely sobering, as they point out a surprisingly high probability of an artifactual association for a typical QSAR investigation. Given the magnitude of a ‘Chance Factor’ effect, drawing independent variables from a pool of hundreds would almost certainly yield equations with significant but totally meaningless correlations.

What’s up with q^2 ?

The leave-one-out (LOO) correlation coefficient, known popularly as q^2 , has received quite a bit of attention in recent years, and the news is not at all good. Investigators typically use the LOO technique in an effort to estimate some degree of QSAR predictivity. The assumption is that, by examining the performance of a QSAR equation derived from all but one molecule, one can obtain a gauge of the ability of the equation to predict properties of new molecules. This is a fallacy. Although q^2 does provide some understanding of the diversity of the molecules under study, it does nothing else. A number of papers have recently appeared to highlight the inadequacies of q^2 [11, 21, 22]. For example, the relationship between q^2 and test set predictions gleaned from 37 3D-QSAR papers published over the past decade (61 models) is plotted in Fig. 9. The graph plainly depicts absolutely no association between q^2 and test set predictivity. In fact, the q^2

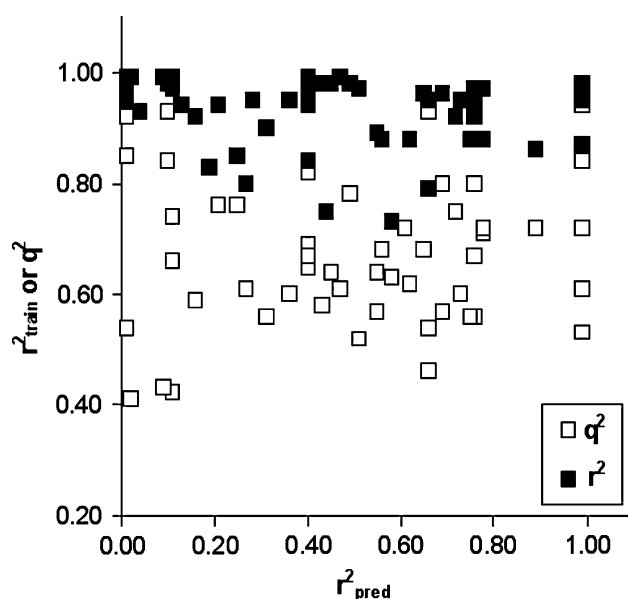


Fig. 9 The relationship between q^2 , test set predictions (r^2_{pred}) and training set r^2 . Data taken from 37 separate literature reports (see Ref. [11])

performance follows training set r^2 . In an in-depth re-examination of the Cramer steroid training and test sets [23], Kubinyi reported a curious relationship between q^2 and r^2_{pred} (test set) using PLS and similarity scores (Fig. 10) wherein most r^2_{pred} values of the test set were found to be better than the q^2 values of the training set. In fact, the best r^2_{pred} values occurred at some sub-optimal q^2 range (0.5–0.7). This curiosity has been referred to as the Kubinyi Paradox [24]. The misuse of q^2 as a measure of predictivity continues to the present day, despite all warnings, tactful and dire, to the contrary.

The QSAR enigma

It has often been said rather sardonically that QSAR works well to highlight SAR trends in a retrospective manner, which is another way of saying that it kicks in nicely once the project is over. Prospective QSAR requires some semblance of predictivity, and this can only happen when a correlation equation is based upon a real world causative mechanism (the assumption is that a causative mechanism will provide for the type of extrapolation necessary for prediction). In a universe of entangled molecular descriptors predictive QSAR remains a considerable challenge. However, the situation is not without hope. We have

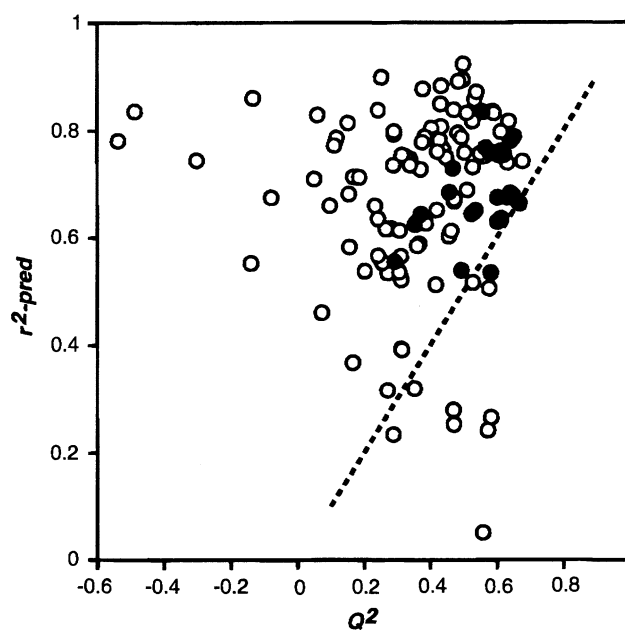


Fig. 10 Relationship between q^2 values of the training set and r^2_{pred} values for the test sets. Twenty-four PLS models are included with one to five latent variables describing the Cramer steroid set activities (figure reprinted from Ref. [22] by permission of the American Chemical Society)

already discussed the dark consequences of a large molecular descriptor space and its inherent pitfalls. Some additional guidelines are discussed herewith.

The general impression is that the success of a QSAR equation is based on its ability to predict the activities of compounds as yet untested, and intuitively, we can guess that this ability rests largely on the choice of the training set. In principle, in order to fairly judge the value of a QSAR, we should limit our choice of test set molecules to those most similar to the training set. However, herein lies an enigma. The dearth of compounds at the beginning of a synthetic program makes it difficult to carry out any such well-reasoned campaign. However, it may be possible to choose synthetic targets early on which could provide compounds embodying extremes of molecular descriptor space historically associated with activity (e.g., hydrophobic, steric and electronic properties). The reality is that the molecules selected for the training set should be representative of each descriptor eventually found to be important ... but we do not know that until later in the synthetic campaign! A practical application may be to incrementally incorporate new information into the QSAR model, re-assess it, and synthesize molecules designed to test the modified model. This process needs to be repeated over and over until a stable QSAR equation surfaces which satisfactorily explains the evolved SAR. At some point the QSAR/synthetic cycle will come to an end because of practical limitations even though the learning cycle that this process represents may be incomplete. This QSAR-guided discovery process does not necessarily require a robust and fully predictive set of correlations. Indeed, it simply requires iteration using meaningful molecular descriptors. Thus, a prospective QSAR, by this definition, serves to identify optimal congeners through guidance rather than global prediction.

Important principles for the development of a sound QSAR model have regularly appeared in the literature [25–32]. The admonitions typically cited include items like (1) choose training set compounds based on relevant descriptors, (2) do not extrapolate beyond the limitations inherent in the training set, (3) avoid improving the correlation beyond the limits of the error in the biological data, (4) take care to avoid mixing classes of molecules (different chemotypes may have similar endpoints but different mechanisms of action), and (5) confirm that the observations reflect a singular biological endpoint (no mixing of mechanism of action; stick to one binding model), to name a few. Point (3), the over-fitting issue, is as interesting as it is common. Frequently QSAR equations are constructed using statistics designed to judge the mathematical significance of adding an independent variable. Following this paradigm often results in correlations with impressive training set r^2 values. The error embedded in the biological

data is often ignored, and the resulting correlation equation will then also model the error, leading to its potential downfall when applied to a test set. Studies designed to determine the impact of error in the biological data have shown that the r^2 performance of a perfect QSAR equation (where all descriptors are relevant and causative) will be significantly limited by error in the data. For example, an observational error of about 2-fold (typical of many biological assays) applied to a dataset of 19 compounds is equivalent to a standard error of 0.2–0.3 log units, which limits the perfect model to a r^2 of 0.77–0.88 [33]. Thus, any QSAR model sporting an unusually robust training set r^2 of 0.8 or greater should be viewed with suspicion.

Additional pitfalls or limitations of QSAR have been recently brought to the attention of the scientific collective. An interesting point recently raised by Maggiora is the likely involvement of “activity cliffs” in the structure–activity “surface,” indicating that the optimized surface is not as smooth as anticipated [34]. Although one can imagine such cliffs associated with the unexpected addition or loss of a hydrogen bond, or an unforeseen steric bump in the binding site, or any number of discontinuous phenomena, it is unlikely that this effect is exclusively responsible for the disappointing lack of predictivity in modern QSAR. The fact that we consistently arrive at wrong models is likely related to the over-arching irrelevant or chance correlation issue raised earlier. As a colleague, Stephen Johnson, has so elegantly pointed out, “Statistics must serve science as a tool; statistics cannot replace scientific rationality, experimental design, and personal observation.”

Alive and well

We have discussed a number of QSAR pitfalls and caveats, and have a general appreciation of the types of misunderstanding and misuse the methodology has had to endure since its inception. Although this essay is not meant to be an exhaustive expose, hopefully it has highlighted how easily things can go awry. We should also be pleased to realize that QSAR is inherently a valuable tool based on sound statistical principles which can, at the very least, retrospectively explain SAR and, at the most, provide synthetic guidance leading to experimentally testable hypotheses. These qualities alone validate QSAR as a viable and important medicinal chemistry tool. Its first cousin, QSPR (quantitative structure–property relationships) has been used successfully for many years, in particular as applied to predicting solubility, LogP and other measurable physical properties. QSAR is commonly and successfully used to optimize process yields, formulations and final product quality. We find its tenets embedded in the form of similarity/diversity metrics

designed to effectively mine databases and devise informative compound libraries. The parameterization of docking/scoring paradigms, particularly those based on calculated estimates of intermolecular interaction energetics, is purely based on the fundamentals of QSAR. Interestingly, challenges persist in developing predictive docking/scoring methods such as these, as recently highlighted in a critical assessment [35], likely because of the descriptor overload issue discussed earlier. A methodology less bloated by descriptors and showing signs of promise as a predictive estimator of binding affinity is LRM (linear response method) [36, 37]. This approach correlates the force field-base estimates of several types of interactions between molecule and binding site through molecular dynamics simulations. A series of related molecules evaluated in this manner provides the descriptors needed for correlation. Thus, QSAR lives on, not only as a stand-alone technique, but even more so in disguised forms within the more popular drug design approaches of the modern era. Correlative thinking has pervaded humankind's existence for eons, evolving from the recognition of danger engendered by the hairy fellow with a rock in his hand to the present day molecular nuance of a well-placed methyl group and its predicted effect on activity. Rebirth gives rise to novel applications of the technique. To paraphrase, "QSAR is dead, QSAR is dead, long live QSAR!"

Acknowledgments Prior to writing this essay I solicited opinions from a number of individuals having had some impact on the general scope and progress of QSAR during the last decade or two. I'd like to think that much of the substance of this perspective was derived both directly and indirectly from the kind comments and advice I received from the following individuals: Marvin Charton, Curt Brennenman, Alex Tropsha, John Block, Dave Stanton, Peter Jurs, Hugo Kubinyi, Richard Lewis, Gerhard Klebe, Gerry Maggiora, Mick Lajiness, Andrew Good, Stephen Johnson and Yvonne Martin. I also thought it might be worthwhile to include some of those comments here (in no particular order):

Regarding QSAR's death: "As Mark Twain once said in response to an article concerning his alleged demise, 'Reports of my death have been greatly exaggerated'."

"QSAR suffers from the extrapolation vs interpolation problem ... by its very nature it is an extrapolative adventure!"

"Without exaggeration at all, I would say that QSAR/QSPR methods are our number-one tools in the tool box."

"Without a (completely independent) test set which has never (!) seen the model, you may cheat yourself and others."

"As a consequence of so many (artificial) parameters and despite the fact that the use of too many parameters has been criticized already thirty years ago, the literature is now spoiled with, most probably, thousands of meaningless chance correlations."

"One of the problems faced in drug design is our lack of knowledge as to which step in the voyage of the drug from point of entry to receptor-drug binding actually determines bioactivity."

"... current descriptor technology is inadequate for capturing the most pertinent effects, and can often cause 'surrogate' descriptors to

appear important in a model, when they actually do not have a physical relationship with the endpoint."

References

1. Hansch C, Fujita T (1964) *J Am Chem Soc* 86:1616
2. Sies H (1988) *Nature* 332:495
3. Diamond G (1988) *Am J Cardiol* 63:392
4. U.S.B.o.t. Census, Washington, DC, 1986
5. Pearson K (1897) *Proc R Soc Lond* 60
6. Stigler S (2005) *Perspect Biol Med* 48:S88
7. Stoupe E, Abramson E, Israelevich P, Sulkes J, Harell D (2007) *Eur J Intern Med* 18:124
8. Barlow RB (1979) *Trends Pharmacol Sci* 1:109
9. Bell PH, Richard J, Roblin O (1942) *J Am Chem Soc* 64:2905
10. Guccione S, Doweiko AM, Chen H, Barreta GU, Balzano F (2000) *J Comput-Aided Mol Des* 14:647
11. Doweiko AM (2004) *J Comput-Aided Mol Des* 18:587
12. Rekker RF (1992) *Quant Struct-Act Relat* 11:195
13. Kubinyi H (2002) *Quant Struct-Act Relat* 21:348
14. Crum-Brown A, Fraser TR (1868–1869) *Trans R Soc Edinburgh* 25:151
15. Richet C (1893) *CR Seances Soc Biol* 9:775
16. Hammett LP (1970) *Physical organic chemistry. Reaction rates, equilibria and mechanism*. McGraw-Hill, New York
17. Free SM, Wilson JW (1964) *J Med Chem* 7:395
18. Martin YC (1981) *J Med Chem* 24:229
19. Fujita T (1984) *Drug Des: Fact Fantasy? Proc Rhone-Poulenc Round Table Conf 3rd edn*:19
20. Topliss JG, Edwards RP (1979) *J Med Chem* 22:1238
21. Golbraikh A, Tropsha A (2002) *J Mol Graphics Modell* 20:269
22. Kubinyi H, Hamprecht FA, Mietzner T (1998) *J Med Chem* 41:2553
23. Cramer RD, Patterson DE, Bunce JD (1988) *J Am Chem Soc* 110:5959
24. Drie JHV (2003) *Curr Pharmaceut Des* 9:1649
25. Cronin MTD, Shultz TW (2003) *J Mol Struct (Theochem)* 622:39
26. Dearden JC, Cronin MTD (2006) *Smith Williams' Introd Princ Drug Des Action*, 4th edn:185
27. Dunn WJ III (1990) *Drug Discov Technol* 22
28. Sheridan RP, Feuston BP, Maiorov VN, Kearsley SK (2004) *J Chem Inf Comput Sci* 44:1912
29. Sjoestroem M, Eriksson L (1995) *Methods Princ Med Chem* 2:63
30. Walker JD, Dearden JC, Schultz TW, Jaworska J, Comber MHI (2003) *Quant Struct-Act Relat Pollut Prev Toxic Screening Risk Assess Web Appl* 3
31. Walker JD, Jaworska J, Comber MHI, Schultz TW, Dearden JC (2003) *Environ Toxicol Chem* 22:1653
32. Doweiko AM (2006) *Compr Med Chem* 4:575
33. Doweiko AM, Bell AR, Minatelli JA, Relyea DI (1983) *J Med Chem* 26:475
34. Maggiora GM (2006) *J Chem Inf Model* 46:1535
35. Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS (2006) *J Med Chem* 49:5912
36. Aqvist J, Medina C, Samuelsson JE (1994) *Protein Eng* 7:385
37. Jones -Hertzog DK, Jorgensen WL (1997) *J Med Chem* 40:1539