HISTORICAL DEVELOPMENT OF QSAR

Historical Development of QSAR

Over the past two decades, the center of gravity (the intellectual focus) of medicinal chemistry has shifted dramatically from, how to make a molecule, to what molecule to make. The challenge now is the gathering of information to make decisions regarding the use of resources in drug design. The information feeding the drug design effort is increasingly quantitative, building upon recent developments in molecular structure description, combinatorial mathematics, statistics, and computer simulations. Collectively these areas have led to a new paradigm in drug design which has been referred to as QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP (QSAR). It has been nearly 40 years since the QSAR paradigm first found its way into the practice of pharmaceutical chemistry. Crum-Brown and Fraser published equation 1.1 in 1868, which is considered to be the first formulation of a QSAR: the “physiological activity “(Φ) was expressed as a function of the chemical structure C.

$$\Phi = f(C) \quad (1.1)$$

A few decades later Richet, Meyer and Overton independently found linear relationship between lipophilicity expressed as solubility or oil-water partition coefficient and biological effects, like toxicity and narcotic activity. In 1930’s, L.Hammett correlated electronic properties of organic acids and bases with their equilibrium constants and reactivity. Taft devised a way for separating polar, steric, and resonance effects and introducing the first steric parameter. The contributions of Hammett and Taft together laid the mechanistic basis for the development of the QSAR paradigm by Hansch and Fujita.254 They combined hydrophobic constants with Hammett’s electronic constants to yield the linear Hansch equation and its many extended forms.

$$\text{Log } 1/C = a\sigma + b\pi + ck \text{ .... Linear form } (1.2)$$

$$\textbf{Log 1/C = a log P – b (log P)2 + c}\sigma\textbf{ + k ………… Non linear form (1.3)}$$

Where,

C - Concentration required to produce a standard response

Log P - partition coefficient between 1-octanol and water

σ - Hammet substituent parameter

π - Relative hydrophobicity of substituents

a, b, c, k - Model co-efficient

Besides the Hansch approach, other methodologies were also developed to tackle structure activity questions. The Free-Wilson approach addresses structure activity studies in a congeneric series as described in Equation (1.4).

$$\mathbf{BA = \Sigma\ a\ i\ x\ i + u\ (1.4)}$$

Where BA is the biological activity, u is the average contribution of the parent molecule, and a i is the contribution of each structural feature; xi denotes the presence x i = 1 or absence x i = 0 of a particular structural fragment. Limitations in this approach led to the more sophisticated Fujita-Ban equation that used the logarithm of activity, which brought the activity parameter in line with other free energy-related terms.

$$\mathbf{Log\ BA = \Sigma\ G\ i\ X\ i + u\ (1.5)}$$

u is defined as the calculated biological activity value of the unsubstituted parent compound of a particular series. G i represents the biological activity contribution of the substituents, whereas X i is ascribed with a value of one when the substituent is present or zero when it is absent. Variations on this activity based approach have been extended by Klopman et. al. and Enslein et al. Topological methods have also been used to address the relationships between molecular structure and biological activity. The Minimum Topological Difference (MTD) method of Simon and the extensive studies on molecular connectivity by Kier and Hall have contributed to the development of quantitative structure property/activity relationships. Recently, these electro topological indices that encode significant structural information on the topological state of atoms and fragments as well as their valence electron content have been applied to biological and toxicity data. Other recent developments in QSAR include approaches such as HQSAR (Hologram QSAR), Inverse QSAR, and Binary QSAR.

THE BEGINNINGS OF CORRELATION (1816-1900)

Physicochemical Properties: The Periodic Table

So far as is known, the earliest work on property prediction involved the elements, or as many of them as were known in the early nineteenth century. The first to publish on this topic appears to have been Döbereiner, who in a letter to Goethe in 1816 (Kauffman, 1999) mentioned what was to evolve into his *Dreiheit* (rule of triads); "The mineral coelestine [strontium sulphate] shows remarkable relationships: its specific weight is the mean of that of [calcium sulphate] and [barium sulphate], namely (2.95 + 4.47)/2 = 3.71". The rule of triads is in effect a read-across technique (Van Leeuwen, Schultz, Henry, Diderich, & Veith, 2009) whereby a property value of a chemical is predicted from known values of that property from one or more similar chemicals. Nonetheless, it is a valid predictive approach, and thus can be included as an historical use of QSPR.

Some of Döbereiner's publications on the subject (Döbereiner, 1817, 1829) extended his ideas to many other families of elements, and in fact his was the first of numerous attempts to produce a systematic arrangement of the elements. A good and concise summary of those attempts has been given by van Spronsen (1959). The elucidation of patterns of elements was not helped by confusion over exact (or even approximate) values of atomic weights. Cooke (1854) developed a classification based on simple formulae for six different series, several of which were based on the number 8, rather than on triads. Odling (1857) and Dumas (1857) were able to make connections between what were formerly considered independent groups of elements. Zapffe (1969) stated that "it could easily be argued that Dumas was the initiator of the whole Mendeleevian stream of thought. He even went so far as to anticipate Mendeleev on the matter of predicting undiscovered elements where gaps appeared in the relationships". Newlands (1864) propounded his "law of octaves", observing an eightfold regularity among the elements. However, the first person actually to develop a periodic system of the elements was Béguyer de Chancourtois (1862), who formulated a helical pattern of groups of elements. Another significant step was taken by Hinrichs (1867) with his hand-written *Atomechanik* giving a group classification of the elements. In 1869 Hinrichs published details of his classification (Hinrichs, 1869).

It is surprising that, with all that went on in connection with classification of the elements prior to that of Mendeleev, his is the chief, and sometimes the only, name that is recalled. That is not to belittle his achievements, which were outstanding. He formulated the Periodic Law, created his own version of the periodic table of elements, and used it to correct the properties of some already discovered elements and also to predict the properties of eight elements yet to be discovered (Mendeleev, 1869). He is reported (Wikipedia) as saying: "I saw in a dream a table where all elements fell into place as required. Awakening, I immediately wrote it down on a piece of paper, only in one place did a correction later seem necessary."

Other Physicochemical Properties

Kopp (1842) observed regularities in the specific volumes, densities and boiling points of compounds within homologous series. His numbering is hard to follow, because he did not define units, and his atomic weights, whilst no doubt in line with the knowledge of his time, were wildly erroneous; for example, he gave the atomic weight of carbon as 75.854, and he appears to use "atomic weight" also to mean "molecular weight". Concerning boiling points, he reported for example that those of the oxalates are all 106° higher than those of the acetates, and that substitution in any series of a methyl group by ethyl raised the boiling point by 18°. Since he looked only at relatively low molecular weight compounds, he failed to realise that within homologous series the boiling point difference between adjacent homologues (Δbp) decreases as chain length increases.

In a later publication Kopp (1844) gave more details of his calculation of specific volumes. For example, he obtained the specific volumes (units unspecified) at 35°C of carbon (72.10), hydrogen (27.04) and oxygen (54.08), and was then able to calculate the specific volume of diethyl ether as 612.9, which he claimed agreed exactly with the measured value. This early approach is similar to that of McGowan (1986) for the calculation of "characteristic volumes".

Mills (1884a) actually formulated a quantitative structure-property relationship to model the melting and boiling points of a number of homologous series, which took account of the decrease in Δbp with increasing chain length. His generic QSPR was:

$$\text{mp or bp} = \frac{\beta(x-c)}{1+\gamma(x-c)} \tag{2}$$

where x = number of CH_2 groups in the chain, and β, γ and c are constants. Mills devised separate equations for odd- and even-numbered chains, and found that for many series melting and boiling points could be predicted to within 1 or 2 degrees. This can probably be considered to be the first proper QSPR.

Mills (1884b) also attempted to model the periodicity of the elements, using a rather complicated formula. He did not appear to be aware of the work of Mendeleev (1869)!

Biological Activities

Blake (1841) reported the toxic effects of inorganic salts on animals such as dogs and rabbits; he found that within what he termed isomorphous groups of metal salts (which in general resemble the divisions later proposed for the Periodic Table (*vide ultra*)) "the action of substances in the same isomorphous group is a function of the atomic weight (of the metal), the greater the atomic weight the smaller the quantity required to produce the same physiological action". In a later work (Blake, 1884) he expanded his studies to include forty elements, but commented that "with these extended investigations no facts have been met with that were not in accord with the conclusions I then (i.e. in 1841) published".

In a M.D. thesis at the University of Strasbourg, Cros (1863) reported an increase in the toxicity of alcohols with decreasing aqueous solubility, up to a maximal potency, limited presumably by very low solubility of the higher alcohols. Similar variations of toxicity with aqueous solubility were observed by Rabuteau (1870), Dujardin-Beaumetz & Audigé (1876) and Richet (1893), but none cited earlier works (Lipnick, 1990). Richardson (1869) observed that the narcotic effect of primary alcohols in several species increased in proportion to their molecular weight. Brunton (1880) expressed the hope that physicians would soon be able to predict the pharmacological action of any compound from its chemical constitution. It is a sombre thought that even now, 136 years on, that goal has not yet been fully achieved. As pointed out above, Crum Brown and Fraser (1868-1869) showed that the pragmatic approach was to examine how changes in chemical constitution (structure) could produce changes in biological response. Bynum (1970) has reviewed much of the 19th century work on the relationship between chemical structure and pharmacological action, and Parascandola (1971) has given more details of the work of Blake, Richardson, Crum Brown & Fraser, and Brunton. It is pleasing to note that Richardson, Brunton and Fraser were all knighted for their services to chemistry. Walker, Enache and Dearden (2003) have reviewed the early work on structure-toxicity relationships of metals and metal ions.

What those who related toxicity to aqueous solubility and molecular weight failed to recognise is that both properties were surrogates for the partition coefficient. Hansch et al. (1968) showed that aqueous solubility of liquids was inversely related to partition coefficient. For homologous series, with successive members differing by CH_2, molecular weight is directly proportional to the logarithm of the partition coefficient. The significance of partition coefficient cannot be over-emphasised; it is undoubtedly the main factor controlling the transport of a chemical within an organism, as it repeatedly crosses lipid membranes on its "random walk" to a site of action (Hansch, 1969).

The breakthrough – not too strong a word – came with the work of Overton (1897, 1899) and Meyer (1899), working independently. They found that the toxicity of narcotic compounds to tadpoles and other aquatic species could be correlated with their oil-water partition coefficients; olive oil was the main oil that they used. Overton's book, *Studien über die Narkose* (1901) has been translated, with additions, into English by Lipnick (1991).

Although neither Overton nor Meyer produced a correlation equation, Hansch (1969) used some of Overton's tadpole data in a QSAR (equation 3), using the 1-octanol-water partition coefficient (P), whilst Leo et al. (1969) found that the Overton data correlated only poorly with other physicochemical properties (polarisability, molar attraction, parachor and molecular weight).

$$\log 1/C = 0.94 \log P + 0.87 \tag{3}$$

$n = 51$ $r^2 = 0.943$.

where C = concentration to produce narcosis in a given time, and n = number of compounds used to develop the model.

The high coefficient of variation (r^2) is tribute to the accurate work of Overton.

MORE PROGRESS (1901-1962)

The level of QSAR/QSPR type activity in the late 19th century continued into the early 20th century. In 1903 Longinescu derived the following expression for the melting points of organic compounds (Godavarthy, Robinson, & Gasem, 2006):

$$T_m = 8.37D\sqrt{n} \tag{4}$$

where D = density, and n = number of atoms in molecule.

A number of other models for melting point prediction appeared over the years, and these have been discussed by Dearden (1999).

Fühner (1903, 1904) found that the toxicity of alcohols to sea-urchin eggs increased with molecular weight, up to a limit imposed by lack of aqueous solubility. He also pointed out that within homologous series narcotic toxicities increased in a geometric progression: 1, 3, 3^2, 3^3 and so on, suggesting additivity of group contributions.

Using some of Overton's data, Traube (1904) found that the toxicity of 11 simple organic compounds increased with the surface tension of the compounds in aqueous solution, and proposed that this related to the ability to penetrate membranes. Vernon (1911) found that the ability of a series of 10 alcohols to lower tortoise heart rate ranked well with Overton's data for tadpole narcosis by the same alcohols.

Holt (1916) found that the toxicity of volatile organic compounds to the cockroach increased as boiling point increased up to a certain point, beyond which toxicity fell again. A similar result was observed by Moore (1917a), who also examined the relationship between toxicity to insects and volatility (Moore, 1917b) and, for liquid insecticides, their surface tension (Moore, 1918). Warburg (1921) investigated narcosis of avian blood cells by simple organic compounds. He considered that narcotics acted by surface properties, and, like Traube before him (Traube, 1904), found that narcosis appeared to be related to surface tension.

Numerous publications in the 1920s and 1930s sought to relate biological activity to molecular weight or aqueous solubility. They include the work of Tilley & Schaffer (1926, 1928), Dohme et al. (1926), and Klarmann et al. (1931). The latter two found activity to pass through a maximum as molecular weight increased. But it is remarkable that, despite the work of Overton and Meyer at the end of the 19th century (*vide ultra*), there appears to be not a single publication in the 1920s and 1930s that used oil-water partition coefficient as a key descriptor in modelling biological activities.

Much of the work cited above was quantified by Ferguson (1939), who showed that the toxic concentrations (C_T) of a series of compounds could be predicted from their aqueous solubility or vapour pressure using equation 5:

$$C_T = kA^{1/n} \qquad (5)$$

where k and n are constants, with $n > 1$, and A = solubility or vapour pressure. Later he developed a generalised form of equation 5 in which k can be another physicochemical property such as partition coefficient (Ferguson & Pirie, 1948). Purcell et al. (1973) have pointed out that equation 5 can be transformed logarithmically into the format now widely used in QSAR/QSPR work:

$$\log 1/C = k' + n'.\log A \qquad (6)$$

Ferguson (1939) also pointed out that "physiological effects due to a physical mechanism are generally measured when an equilibrium has been established between the concentration of the toxic substance in the phase, solution or vapour, in which it is applied and the concentration in the phase or surface layer which is the seat of toxic action. Since the chemical potential of the toxic substance must be the same in all phases partaking in the equilibrium, it is suggested that the toxicities of physically toxic substances should be compared, not by the value of the toxic concentration in the external solution or vapour, but by the values of the chemical potentials in these phases. Chemical potentials so determined are identical with the chemical potentials at the actual point of attack within the organism. Using the activity function of G.N. Lewis as the chemical potential, it is found that, when chemical reaction is absent, though diverse chemical compounds exert the same toxic effect on a given organism at widely different concentration, the activities corresponding to these concentrations lie within a relatively narrow range". For example, for a series of 14 narcotic gases and vapours tested on mice, whilst the narcotic concentrations ranged from 0.5% - 100%, the corresponding activities ranged only from 0.01 to 0.07.

Another great step forward was the development of electronic substituent constants by Hammett (1935, 1937), which could be correlated with reaction and equilibrium constants within a chemical series using equation 7:

$$\log (k_X/k_H) = \rho\sigma \qquad (7)$$

where k_H is the reaction rate or equilibrium constant of the parent compound and k_X is that of a derivative with substituent X; ρ is the series constant, and σ is what is now called the Hammett substituent constant. Hammett's original work involved the acid dissociation constants of a series of benzoic acids, for which he set ρ equal to unity. For other series and other equilibria or rate constants, ρ has different values. Hammett's work set in train a new field of study called linear free energy relationships (LFER). This name derives from the van't Hoff isotherm, which shows that the free energy change in a process is directly proportional to the logarithm of the equilibrium or rate constant of the process:

$$\Delta G = -2.303RT\log K \qquad (8)$$

where ΔG = Gibbs free energy, R = universal gas constant, T = absolute temperature, and K = an equilibrium or rate constant. Following Hammett, much work was done in modelling physical and chemical processes using the Hammett substituent constant (Wells, 1968). However, it was soon recognised that σ has two components, resonance (R) and inductive (I) effects; furthermore, it is position-dependent. Hansch and Leo (1979) have discussed these and other effects in some detail, and give an example of the modelling of pKa values of benzoic acids:

$$pKa = 5.76 - 1.64\,\sigma_I^m - 0.34\,\sigma_R^m - 1.76\,\sigma_I^p - 1.15\,\sigma_R^p \qquad (9)$$

$n = 27$ $r^2 = 0.984$ $s = 0.083$

where superscripts m and p refer to meta- and para-substitution. Hansch and Leo (1979) also included in their book many values of a number of types of Hammett constants.

Hammett constants model electronic effects, and hence Taft (1952) devised a steric constant, E_s, in order to incorporate steric effects into LFERs.

Although the Hammett substituent constant was developed and used successfully for correlating physicochemical properties, it is not surprising that attempts were made to examine whether it could be used in the correlation of biological activities. For example, Hansen (1962) reported the toxicities and σ constants of benzoic acids to mosquito larvae, from which Kubinyi (1993) developed the following QSAR:

$$\log 1/C = 1.787 + 1.454\,\sigma \qquad (10)$$

$n = 13$ $r^2 = 0.843$ $s = 0.243$

Zahradnik and Chvapil (1960) also attempted to correlate various biological activities with Hammett and Taft constants, with only limited success. Jaffé (1953) commented that he had not found any series of compounds whose biological activity could be described by Hammett constants.

A number of boiling point correlations published in the first half of the 20th century have been listed by Egloff et al. (1940); they also published their own QSPR for the prediction of boiling points (T) of normal alkanes:

$$T = 745.42 \log (n + 4.4) - 416.31 \tag{11}$$

where n = number of carbon atoms. Other than for methane, the prediction errors averaged 0.40°. Surprisingly, Egloff et al. (1940) did not mention the excellent correlations of Mills (1884a).

A very important contribution, almost totally ignored, was that of Bell and Roblin (1942), who investigated the bacteriostatic activities of a series of sulfanilamides. They found a biphasic relationship between activity and pKa, and developed a QSAR to model that relationship:

$$\log (0.001/C_R) = 3.23 + 0.661 \log x + 0.339 \log (1 - x) \tag{12}$$

where C_R = minimum molar concentration to exhibit a given level of bacteriostatic activity, x = fraction of the total concentration in ionic form.

Several things are especially noteworthy about this work. Firstly, a good biphasic correlation was reported. Secondly, equation 12 is very similar to the modern format of QSAR equations, including the use of logarithmic functions, the use of reciprocal concentrations and the use of molar concentrations. Thirdly, the title of the paper includes the phrase "relation of structure to activity", which may be the origin of the now widely used term "structure-activity relationship". Equation 12 is quite probably the very first biological QSAR.

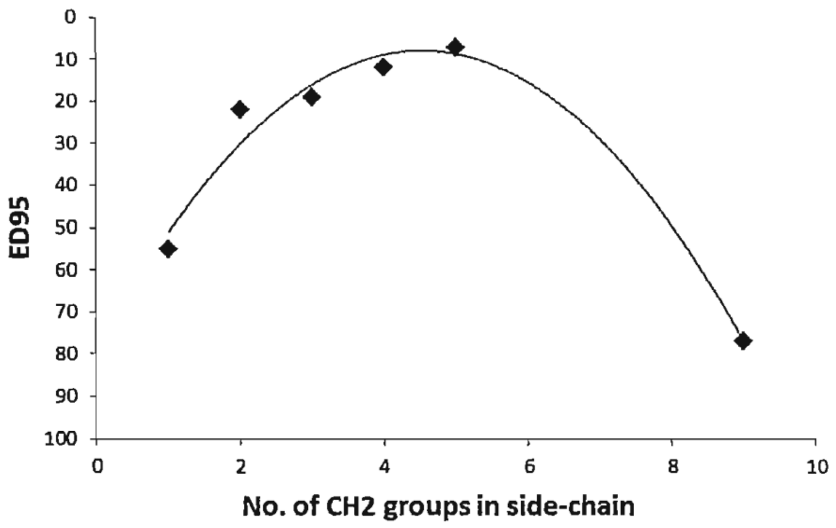
Fieser and Richardson (1948) found that the activities of several series of naphthoquinone antimalarials in ducks varied parabolically with the number of carbon atoms in the side-chain (see Figure 1). Unfortunately they gave no mechanistic interpretation of their observations, although in a later publication (Fieser, Ettlinger, & Fawaz, 1948) they reported that the values of log P of the most active members of eight series of naphthoquinones were all in the range 10 – 12.

A quadratic equation fitting the data shown in Figure 1 has good statistics ($r^2 = 0.974$); when the data are transformed to the logarithm of molar concentrations, as is the correct QSAR practice, the correlation is still good ($r^2 = 0.952$).

McGowan (1952) devised a QSAR for the toxicities of diverse narcotic compounds that gave good correlations:

$$\log C = S + (4 \times 10^{23})\, p - 0.036\, P_c + 7.7 \tag{13}$$

Figure 1. Anti-malarial activity of cyclohexylalkylnaphthoquinones against P. lophurae in ducks (after Fieser & Richardson, 1948)



where S = sensitivity factor based on varying sensitivity of different biological systems, p = polarisability, and P_c = parachor. It should be noted that both polarisability and parachor have units of molecular volume, so in effect McGowan was correlating narcotic toxicity with molecular size, as had many others before him.

Rather than trying to model biological activities using Hammett constants, Bruice et al. (1956) developed empirical substituent constants based on the contribution of each substituent to thyromimetic activity in both amphibians and mammals, and obtained good correlations. The problems with this type of analysis are that new constants have to be derived for each type of biological activity, that the contribution of each substituent is fixed, and that mechanistic interpretation is difficult.

1962 – PRESENT

It was not until 1962 – 65 years after Overton (1897) first showed that oil-water partitioning was the key process in narcosis – that the second major breakthrough occurred. Corwin Hansch, of Pomona College, California, and co-workers (Hansch, Maloney, Fujita, & Muir, 1962) showed that, for a series of phenoxyacetic acid plant growth regulators, the activity in *Avena* (oats) could be correlated with both the Hammett substituent constant σ and a hydrophobic substituent constant π:

$$\log 1/C = 3.36 + 4.08\,\pi - 2.14\,\pi^2 + 2.78\,\sigma \qquad (14)$$

where C = concentration to produce a specified growth in *Avena*.

This seminal work was outstanding in several ways:

1. It showed that complex biological effects, and not simply narcosis, could be modelled by physicochemical descriptors;
2. It showed that more than one type of descriptor was sometimes required to obtain a good model of biological activity;
3. It showed that biological activity can be modelled by a biphasic correlation. As pointed out above, others (Dohme, Cox & Miller, 1926; Klarmann, Gatyas & Shternov, 1931; Bell & Roblin, 1942; Fieser & Richardson, 1948) had found that a plot of biological activity against a physicochemical descriptor often rose to a maximum and then fell again, but no-one had previously represented that by a quadratic equation;

4. It pioneered the use of 1-octanol as a surrogate for lipid. Whilst the analogy is not perfect, both 1-octanol and lipids typically have a polar end-group and a hydrocarbon chain, and octanol-water π and log P values have been found to be valuable descriptors in very many QSAR correlations. Other pinguid phases were later proposed (Leahy, Morris, Taylor, & Wait, 1992), but 1-octanol still remains the most widely used;
5. It pioneered the use of the hydrophobic substituent constant, π, which is defined for a substituent X on a molecule MX as $\pi_X = \log(P_{MX}/P_{MH})$ or $\log P_{MX} - \log P_{MH}$, where P is the octanol-water partition coefficient. For example, the log P of toluene is 2.68, and that of benzene is 2.13, so π_{CH_3} is 2.68 – 21.3 = 0.55.This in effect means that hydrophobicity is, to a first approximation at least, an additive property.

However, it is not quite correct to say that benzene + methyl = toluene; it is phenyl (C_6H_5) + methyl that equals toluene. Hence the calculation of log P of toluene as log $P_{benzene} + \pi_{CH3}$ is wrong by π_H. This is a negligible difference, but for a long-chain compound the error becomes significant. This was pointed out by Nys and Rekker (1973), who introduced a different method of calculating log P values. By analysing the log P values of a large number of compounds, they were able to calculate the contribution of molecular fragments (hydrophobic fragmental constants) to log P. This method avoided the problem found with the hydrophobic substituent constant, as was acknowledged by Leo et al. (1975), who introduced their own fragmental constants, obtained from the careful measurement of log P values of very small molecules such as hydrogen and methane. This work formed the basis for the ClogP software for the calculation of log P.

Hansch and Fujita (1964) expanded their studies by showing that their approach worked with, for example, the toxicity of phenols to bacteria, the toxicity of benzoic acids to mosquito larvae, and the toxicity of diethyl phenyl phosphates to houseflies. Since then the Hansch group have published many papers covering a wide range of chemicals and species, and Hansch (1969) has presented a succinct account of what has come to be called Hansch analysis (Kubinyi, 1993). Hansch and Dunn (1972) and Hansch and Clayton (1973) reported numerous examples of biological activities correlated with log P using rectilinear and parabolic models respectively. Others (van de Waterbeemd, 1992; Kubinyi, 2002; Selassie, 2003; Martin, 2012) have reviewed the QSAR contributions of Hansch and his co-workers. It is not for nothing that Corwin Hansch has been called the father of QSAR (Martin & Stouch, 2011).

Contemporaneously with the ρ-σ-π approach of Hansch and Fujita (1964), Free and Wilson (1964) developed an approach similar to that of Bruice et al (1956), using indicator variables for the presence of structural features different from those of a chosen reference compound. These were correlated with the relevant biological activity data, and the regression coefficients generated represent the contributions of each structural feature to the model. Kubinyi (1993) has commented that the Free-Wilson approach is difficult to use, and the Fujita and Ban modification (1971) offers important advantages. However, the approach is little used today. Kubinyi (1993) has given a detailed analysis of the Free-Wilson and Fujita-Ban approaches.

Most molecules are flexible, and hence are able to adopt a number of conformations. When a chemical binds to a receptor site in an organism, it does so in what can be considered a preferred conformation, which is usually a low-energy conformation. Kier (1967) utilized extended Hückel theory molecular orbital calculations to obtain preferred conformations, and found good agreement with conformations obtained from crystal X-ray analysis. He later reported many other such calculations, for example for thyrotropin-releasing hormone (George & Kier, 1973). Since then many other workers have utilized M.O. theory to generate preferred conformations for QSAR modeling (Martin, 2002).

In the 1970s and 1980s, a story was circulating in QSAR circles that a car had been seen on a U.S. highway with a bumper sticker saying: "Corwin Hansch walks on octanol". At the 1988 European QSAR Symposium, held in Interlaken, Switzerland, when Hansch was 70 years old, lapel badges were given to all delegates with those words on (see Figure 3).

Many scientists found it hard to accept that something as complex as, for example, anti-inflammatory activity of drugs, or insecticidal activity, can be modelled by a few physicochemical and/or structural properties of chemicals. In fact, what QSAR does is to model the change of biological activity with change of chemical structure, as Crum Brown and Fraser (1868-1869) pointed out many years ago. Most biological processes are multi-step, but generally one of those steps will be rate-limiting, and so the whole process can be modelled with (often simple) chemical properties. An example is the anti-inflammatory activity in the rat of a series of aspirin derivatives (Dearden & George, 1979), which is modelled by the 1-octanol-aqueous buffer (pH1) partition coefficient alone, as Figure 4 shows.

Figure 2. Professor Corwin Hansch, father of QSAR: 6 October 1918 – 8 May 2011 (photograph courtesy of the Department of Chemistry, Pomona College, Pomona, CA)

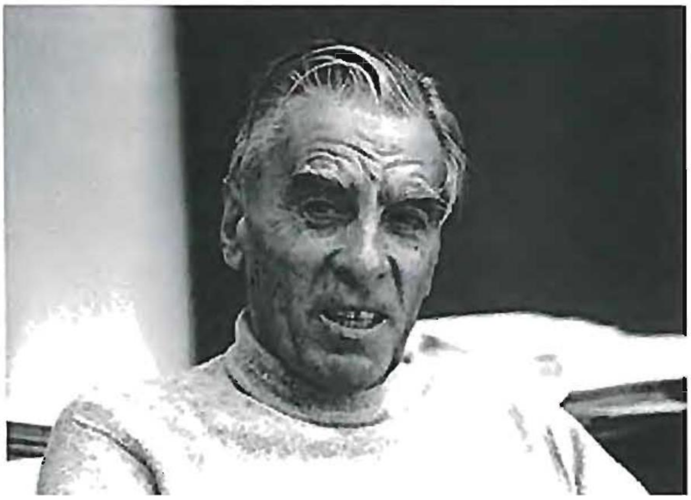
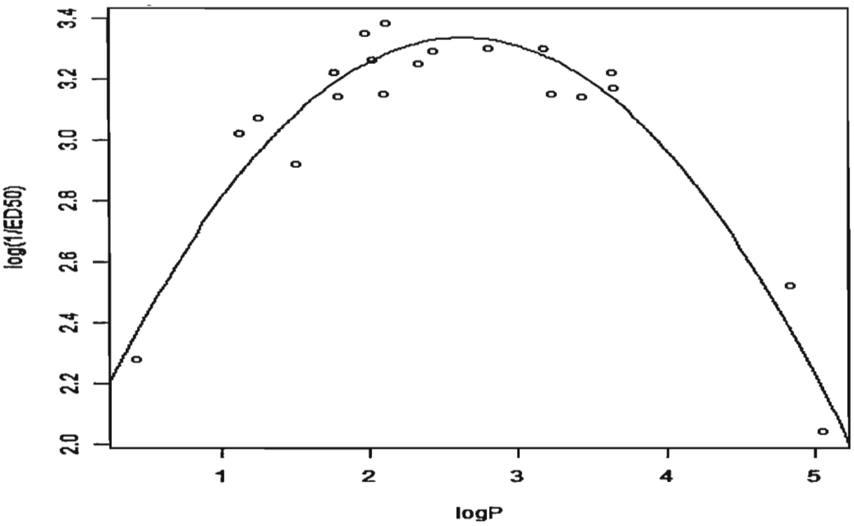


Figure 3. Lapel badge given to delegates attending the 1988 European QSAR Symposium

Figure 4. Anti-inflammatory activity of aspirin derivatives vs. log P



The QSAR for the anti-inflammatory activity is:

$$\log (1/ED_{50}) = 1.958 + 1.029 \log P - 0.195 (\log P)^2 \tag{15}$$

$n = 20$ $r^2 = 0.904$ $s = 0.118$

In the 1970s, the pharmaceutical industry was starting to use QSAR in drug design to a significant extent, and that has carried on (Perkins, Fang, Tong, & Welsh, 2003). There is, however, a story (possibly apocryphal) that medicinal chemists, on finding that increasing the hydrophobicity of drug candidates did not always increase potency, would cry despairingly: "Methyl, ethyl, butyl, futile".

Others were sceptical of the fact that some QSAR models incorporated as many as five or six descriptors, and would claim: "Give me five descriptors and I'll draw you an elephant!" This stimulated Wei (1975) to see how many descriptors actually were needed to fit an elephant. He found that 30 descriptors were required before a shape that could be described as elephant-like was obtained; five descriptors gave merely an egg shape.

With the increasing interest in QSAR in the 1970s, two series of QSAR conferences were initiated. The first was the European Symposia on QSAR, started by Miloň Tichý in Prague in 1973; they were initially held triennially, but from 1982 have been held biennially. Table 1 gives details of those held to date, and the proceedings arising from them.

The QSAR for the anti-inflammatory activity is:

$$\log (1/ED_{50}) = 1.958 + 1.029 \log P - 0.195 (\log P)^2 \tag{15}$$

$n = 20$ $r^2 = 0.904$ $s = 0.118$

In the 1970s, the pharmaceutical industry was starting to use QSAR in drug design to a significant extent, and that has carried on (Perkins, Fang, Tong, & Welsh, 2003). There is, however, a story (possibly apocryphal) that medicinal chemists, on finding that increasing the hydrophobicity of drug candidates did not always increase potency, would cry despairingly: "Methyl, ethyl, butyl, futile".

Others were sceptical of the fact that some QSAR models incorporated as many as five or six descriptors, and would claim: "Give me five descriptors and I'll draw you an elephant!" This stimulated Wei (1975) to see how many descriptors actually were needed to fit an elephant. He found that 30 descriptors were required before a shape that could be described as elephant-like was obtained; five descriptors gave merely an egg shape.

With the increasing interest in QSAR in the 1970s, two series of QSAR conferences were initiated. The first was the European Symposia on QSAR, started by Miloň Tichý in Prague in 1973; they were initially held triennially, but from 1982 have been held biennially.

REFERENCES

1. Roy K, Kar S, Das RN (2015). "Chapter 1.2: What is QSAR? Definitions and Formulism". A primer on QSAR/QSPR modeling: Fundamental Concepts. New York: Springer-Verlag Inc. pp. 2–6. ISBN 978-3-319-17281-1.

2. Qspr/Qsar Studies by Molecular Descriptors UK ed. Edition by Mircea V Diudea

3. Roy K, Kar S, Das RN (2015). "Chapter 1.2: What is QSAR? Definitions and Formulism". A primer on QSAR/QSPR modeling: Fundamental Concepts. New York: Springer-Verlag Inc. pp. 2–6. ISBN 978-3-319-17281-1.

4. Todeschini, Roberto; Consonni, Viviana. Molecular Descriptors for Chemoinformatics. Wiley. ISBN 978-3-527-31852-0.

5. Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens 1st Edition by Romualdo Benigni

6. A Practical Guide to Rational Drug Design 1st Edition, Kindle Edition by Sun Hongmao