Rigorous analysis and fine-tuning of independent variables has led to an expansion in development of molecular and atom-based descriptors, as well as descriptors derived from quantum chemical calculations and spectroscopy (2). The improvement in high-throughput screening procedures allows for rapid screening of large numbers of compounds under similar test conditions and thus minimizes the risk of combining variable test data from many sources.

The formulation of thousands of equations using QSAR methodology attests to a validation of its concepts and its utility in the elucidation of the mechanism of action of drugs at the molecular level and a more complete understanding of physicochemical phenomena such as hydrophobicity. It is now possible not only to develop a model for a system but also to compare models from a biological database and to draw analogies with models from a physical organic database (3). This process is dubbed ***model mining*** and it provides a sophisticated approach to the study of chemical-biological interactions. QSAR has clearly matured, although it still has a way to go. The previous review by Kubinyi has relevant sections covering portions of this chapter as well as an extensive bibliography recommended for a more complete overview (4).

1.1 Historical Development of QSAR

More than a century ago, Crum-Brown and Fraser expressed the idea that the physiological action of a substance was a function of its chemical composition and constitution (5). A few decades later, in 1893, Richet showed that the cytotoxicities of a diverse set of simple organic molecules were inversely related to their corresponding water solubilities (6). At the turn of the 20th century, Meyer and Overton independently suggested that the narcotic (depressant) action of a group of organic compounds paralleled their olive oil/water partition coefficients (7, 8). In 1939 Ferguson introduced a thermodynamic generalization to the correlation of depressant action with the relative saturation of volatile compounds in the vehicle in which they were administered (9). The extensive work of Albert, and Bell and Roblin established the importance of ionization of bases and weak acids in bacteriostatic activity (10–12). Meanwhile on the physical organic front, great strides were being made in the delineation of substituent effects on organic reactions, led by the seminal work of Hammett, which gave rise to the "sigma-rho" culture (13, 14). Taft devised a way for separating polar, steric, and resonance effects and introducing the first steric parameter, E_S (15). The contributions of Hammett and Taft together laid the mechanistic basis for the development of the QSAR paradigm by Hansch and Fujita. In 1962 Hansch and Muir published their brilliant study on the structure-activity relationships of plant growth regulators and their dependency on Hammett constants and hydrophobicity (16). Using the octanol/water system, a whole series of partition coefficients were measured, and thus a new hydrophobic scale was introduced (17). The parameter **a,** which is the relative hydrophobicity of a substituent, was defined in a manner analogous to the definition of sigma (18).

$$\pi_X = \log P_X - \log P_H \tag{1.1}$$

P_X and P_H represent the partition coefficients of a derivative and the parent molecule, respectively. Fujita and Hansch then combined these hydrophobic constants with Hammett's electronic constants to yield the linear Hansch equation and its many extended forms (19).

$$\text{Log } 1/C = a\sigma + b\pi + ck \tag{1.2}$$

Hundreds of equations later, the failure of linear equations in cases with extended hydrophobicity ranges led to the development of the Hansch parabolic equation (20):

$$\begin{aligned}\text{Log } 1/C = a \,.\, &\log \text{P} \\ &- b(\log P)^2 + c\sigma + k\end{aligned} \tag{1.3}$$

The delineation of these models led to explosive development in QSAR analysis and related approaches. The Kubinyi bilinear model is a refinement of the parabolic model and, in many cases, it has proved to be superior (21).