

may be attributed to inaccuracies in the testing procedure (usually dilution errors) or unusual behavior. They often provide valuable information in terms of the mechanistic interpretation of a QSAR model. They could be participating in some intermolecular interaction that is not available to other members of the data set or have a drastic change in mechanism.

2.3 Compound Selection

In setting up to run a QSAR analysis, compound selection is an important angle that needs to be addressed. One of the earliest manual methods was an approach devised by Craig, which involves two-dimensional plots of important physicochemical properties. Care is taken to select substituents from all four quadrants of the plot (63). The Topliss operational scheme allows one to start with two compounds and construct a potency tree that grows branches as the substituent set is expanded in a **stepwise** fashion (64). Topliss later proposed a batchwise scheme including certain substituents such as the **3,4-Cl₂**, **4-Cl**, **4-CH₃**, **4-OCH₃**, and **4-H** analogs (65). Other methods of manual substituent selection include the Fibonacci search method, sequential simplex strategy, and parameter focusing by Magee (66–68).

One of the earliest computer-based and statistical selection methods, cluster analysis was devised by Hansch to accelerate the process and diversity of the substituents (1). Newer methodologies include D-optimal designs, which focus on the use of $\det(X'X)$, the variance-covariance matrix. The determinant of this matrix yields a single number, which is maximized for compounds expressing maximum variance and minimum covariance (69–71). A combination of fractional factorial design in tandem with a principal property approach has proven useful in QSAR (72). Extensions of this approach using multivariate design have shown promise in environmental QSAR with nonspecific responses, where the clusters overlap and a cluster-based design approach has to be used (73). With strongly clustered data containing several classes of compounds, a new strategy involving local multivariate designs within each cluster is described. The chosen compounds from the local

designs are grouped together in the overall training set that is representative of all clusters (74).

3 PARAMETERS USED IN QSAR

3.1 Electronic Parameters

Parameters are of critical importance in determining the types of intermolecular forces that underly drug-receptor interactions. The three major types of parameters that were initially suggested and still hold sway are electronic, hydrophobic, and steric in nature (20, 75). Extensive studies using electronic parameters reveal that electronic attributes of molecules are intimately related to their chemical reactivities and biological activities. A search of a computerized QSAR database reveals the following: the common Hammett constants (σ , σ^+ , σ^-) account for 7000/18500 equations in the Physical organic chemistry (PHYS) database and nearly 1600/8000 in the Biology (BIO) database, whereas quantum chemical indices such as HOMO, LUMO, BDE, and polarizability appear in 100 equations in the BIO database (76).

The extent to which a given reaction responds to electronic perturbation constitutes a measure of the electronic demands of that reaction, which is determined by its mechanism. The introduction of substituent groups into the framework and the subsequent alteration of reaction rates helps delineate the overall mechanism of reaction. Early work examining the electronic role of substituents on rate constants was first tackled by Burckhardt and firmly established by Hammett (13, 14, 77, 78). Hammett employed, as a model reaction, the ionization in water of substituted benzoic acids and determined their equilibrium constants K_a . See Equation 1.28. This led to an operational definition of σ , the substituent constant. It is a measure of the size of the electronic effect for a given substituent and represents a measure of electronic charge distribution in the benzene nucleus.

$$\sigma_X = \log K_X - \log K_H \quad \text{or} \quad (1.29)$$

$$\log(K_X/K_H) = -\rho K_X + \rho K_H$$

Electron-withdrawing substituents are thus