Table 1.2 Types of Biological Data Utilized in QSAR Analysis

Source of Activity	Biological Parameters
1. Isolated receptors	
Rate constants	Log k_{cat}; Log k_{uncat}; Log k
Michaelis–Menten constants	Log $1/K_m$
Inhibition constants	Log $1/K_i$
Affinity data	pA_2; pA_1
2. Cellular systems	
Inhibition constants	Log $1/IC_{50}$
Cross resistance	Log *CR*
In vitro biological data	Log $1/C$
Mutagenicity states	Log TA_{98}
3. "In vivo" systems	
Biocencentration factor	Log *BCF*
In vivo reaction rates	Log I (Induction)
Pharmacodynamic rates	Log T (total clearance)

It is also important to design a set of molecules that will yield a range of values in terms of biological activities. It is understandable that most medicinal chemists are reluctant to synthesize molecules with poor activity, even though these data points are important in developing a meaningful QSAR. Generally, the larger the range (>2 log units) in activity, the easier it is to generate a predictive QSAR. This kind of equation is more forgiving in terms of errors of measurement. A narrow range in biological activity is less forgiving in terms of accuracy of data. Another factor that merits consideration is the time structure. Should a particular reading be taken after 48 or 72 h? Knowledge of cell cycles in cellular systems or biorhythms in animals would be advantageous.

Each single step of drug transport, binding, and metabolism involves some form of partitioning between an aqueous compartment and a nonaqueous phase, which could be a membrane, serum protein, receptor, or enzyme. In the case of isolated receptors, the endpoint is clear-cut and the critical step is evident. But in more complex systems, such as cellular systems or whole animals, many localized steps could be involved in the random-walk process and the eventual interaction with a target. Usually the observed biological activity is reflective of the slow step or the rate-determining step.

To determine a defined biological response (e.g., IC_{50}), a dose-response curve is first established. Usually six to eight concentrations are tested to yield percentages of activity or inhibition between 20 and 80%, the linear portion of the curve. Using the curves, the dose responsible for an established effect can easily be determined. This procedure is meaningful if, at the time the response is measured, the system is at equilibrium, or at least under steady-state conditions.

Other approaches have been used to apply the additivity concept and ascertain the binding energy contributions of various substituent (R) groups. Fersht et al. have measured the binding energies of various alkyl groups to aminoacyl-tRNA synthetases (54). Thus the AG values for methyl, ethyl, isopropyl, and thio substituents were determined to be 3.2, 6.5, 9.6, and 5.4 kcal/mol, respectively.

An alternative, generalized approach to determining the energies of various drug-receptor interactions was developed by Andrews et al. (55), who statistically examined the drug-receptor interactions of a diverse set of molecules in aqueous solution. Using Equation 1.9, a relationship was established between AG and E_X (intrinsic binding energy), E_{DOF} (energy of average entropy loss), and the AS,, (energy of rotational and translational entropy loss).

$$\Delta G = T\,\Delta S_{r,t} + n_{DOF}E_{DOF} + n_X E_X \quad (1.9)$$

E_X denotes the sum of the intrinsic binding energy of each functional group of which n_X are present in each drug in the set. Using Equation 1.9, the average binding energies for various functional groups were calculated. These energies followed a particular trend with charged groups showing stronger interactions and nonpolar entities, such as sp^2, sp^3 carbons, contributing very little. The applicability of this approach to specific drug-receptor interactions remains to be seen.

2.2 Statistical Methods: Linear Regression Analysis

The most widely used mathematical technique in QSAR analysis is multiple regression