FUNDAMENTAL ASPECTS OF QUANTITATIVE STRUCTURE-REACTIVITY RELATIONSHIPS

FRANK H. QUINA AND ERICK L. BASTOS

Department of Fundamental Chemistry, Institute of Chemistry, University of São Paulo, São Paulo, SP, Brazil

1 INTRODUCTION AND SCOPE

A basic tenet of chemistry is that molecular structure determines properties of the molecule such as intrinsic stability, reactivity, and solubility. Changes of these properties due to variations in structure are thus in principle both predictable and comprehensible. Classical thermodynamics provides the framework within which the chemist can link structure and stability and unambiguously predict the direction of spontaneous change. Molecular structure-related factors that can affect reactivity include electronic effects (field and resonance interactions of substituents, electrostatic attraction and repulsion, electrophilicity, nucleophilicity, dipolarity, and polarizability), synchrony (or the lack thereof) between bond breaking and bond making, steric effects, hydrogen bonding, solvation, conformation, and mass transport, all of which can involve both enthalpic and entropic components [1–3].

The fundamental goal of the quantitative structure–reactivity relationship (QSRR) approach is to employ sets of parameters or descriptors to correlate molecular structure with one or more properties related to reactivity, preferably via a linear or additive relationship. The requirement that the property of interest be proportional to Gibbs energy is an intuitively reasonable choice. Since standard free energies are additive and path independent, we can devise hypothetical paths from reagent to product that allow us to separate, at least conceptually, the contributions of the individual factors that make up the overall Gibbs energy change. Similarly, in the treatment of reactivity, the Gibbs energy of activation can also be considered to consist of individual additive Gibbs-energy-related contributions. For this purpose, however, one must have in hand an arsenal of appropriate

parameters or descriptors that provide a realistic measure of the potential strength of each of these individual contributions.

This chapter begins with a brief consideration of the relationship between equilibrium constants and Gibbs energy and the connection between rate constants for chemical reactions and the extra-thermodynamic Gibbs energy of activation (i.e., the activation energy), followed by an exemplification of the link between these that is implied by the Hammond Postulate and Marcus theory. This is then followed by a consideration of descriptors, beginning with those for substituent effects on reactivity. Substituent parameters (or constants) encode the relative ability of a given substituent to stabilize or destabilize charge at the reaction center. As a result, they reflect the decrease or increase in the Gibbs energy of activation of the rate-determining step of the reaction in response to changes in electron density at the reaction center. In addition to these electronic effects of substituents, the steric effects of substituents play an important role in determining reactivity and require other sets of descriptors.

The complementary concepts of nucleophilicity and electrophilicity are essential for an understanding of electronic contributions to the reactivity between different molecules, and much progress has been made in developing scales of nucleophilicity and electrophilicity. Nucleophilicity, like most other electronic effects, is subject to the influence of solvation. Solvation can be treated in terms of descriptors that scale with solvophilicity or solvophobicity of the solute or parts of the solute such as a substituent. Since solvation is the combined result of complementary interactions between solute and solvent, sets of solvent descriptors that factor these into specific and nonspecific interactions have been successfully used to analyze the reactivity of solutes in pure solvents and, to a more limited extent, in solvent mixtures. The corresponding sets of solute descriptors provide a more detailed, molecular-level understanding of the solvent–solute interactions for a given solvent or solvent class.

Finally, quantitative structure–activity relationships (QSARs) are briefly considered and some of the more fundamental aspects of the construction of statistically valid and chemically meaningful linear free energy relationships (LFERs; the alternative suggested by IUPAC [4], that is, linear Gibbs energy relationships, is rarely used).

2 EQUILIBRIUM CONSTANTS, RATE CONSTANTS, AND REACTION PROFILES

A convenient starting point for the consideration of QSRRs is the diagram shown in Figure 1, illustrating the Gibbs energy pathway from reagents to products for three hypothetical elementary reactions. In this diagram, the x-axis represents P_{Rxn}, the "progress" or "extent" of the reaction, zero being the reagents prior to reaction and unity the products at the end of the reaction. The standard Gibbs energy difference between the reagents and products, ΔG^0, is related to the corresponding differences in enthalpy (ΔH^0) and entropy (ΔS^0) and the equilibrium constant (K_{eq}) between reagents and products, as indicated in Equation 1:

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 = -RT \ln K_{eq} \tag{1}$$

where R is the ideal gas constant and T the absolute temperature. The transition state for the elementary reaction occurs at an intermediate value of P_{Rxn} along the reaction profile at a Gibbs energy value $\Delta^{\ddagger}G = \Delta^{\ddagger}H - T\Delta^{\ddagger}S$ above that of the reagents, defining
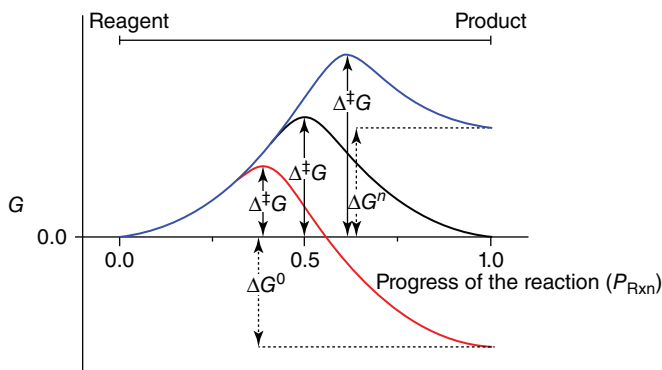
Figure 1   Schematic profiles of the variation of the Gibbs energy as a function of the progress of reaction (P_{Rxn}) for endergonic, isoergonic, and exergonic elementary reactions.

the Gibbs energy barrier to reaction (the notation employed for the extra-thermodynamic activation parameters is that recommended by IUPAC [4]). In Figure 1, as expected from the Hammond–Leffler Postulate [5–7], the transition state for the fairly exergonic reaction is closer to the reagent than to the product, that is, is more reagent-like than product-like, while that for the fairly endergonic reaction is more product-like. Thus, unlike the alternative representation of the reaction profile in terms of the variation of the potential energy along an (often ambiguous or poorly specified) "reaction coordinate," the profile of Figure 1 intrinsically incorporates entropic effects on the reaction rate and unambiguously defines the lowest Gibbs energy reaction path interconnecting reagents and products.

The desired connection to experiment was provided in the 1930s by Eyring's Absolute Reaction Rate Theory, which relates the rate constant for reaction, k_{Rxn}, to the Gibbs energy of activation ($\Delta^{\ddagger}G$) via Equation 2 [8]:

$$k_{Rxn} = \frac{k_B T}{h}(C^0)^{1-n}e^{-\Delta^{\ddagger}G/RT} = \frac{k_B T}{h}(C^0)^{1-n}e^{\Delta^{\ddagger}S/R}e^{-\Delta^{\ddagger}H/RT} \tag{2}$$

where k_B is the Boltzmann constant, h Plank's constant, n the order of the reaction, and C^0 the standard-state concentration (taken to be 1 mol dm^{-3}), which ensures that k_{Rxn} has the appropriate dimensions. Using the relationship $E_a = \Delta^{\ddagger}H + RT$, where E_a is the activation energy, Equation 2 can be cast into a form analogous to that of the empirical Arrhenius equation, as shown in Equation 3:

$$k_{Rxn} = \frac{k_B T \cdot e^1}{h}(C^0)^{1-n}e^{\Delta^{\ddagger}S/R}e^{-E_a/RT} = Ae^{-E_a/RT} \tag{3}$$

allowing experimental estimation of the entropy of activation ($\Delta^{\ddagger}S$) from the Arrhenius pre-exponential A, the enthalpy of activation from the value of E_a, and hence the Gibbs energy of activation for the reaction.

3   MARCUS THEORY

In the development of his classical theory of outer-sphere electron-transfer reactions, Marcus approximated the Gibbs energy profile along P_{Rxn} in Figure 1 in terms of two parabolas.

Marcus theory provided a much clearer understanding of the intrinsic connection between reactivity, which is dependent on $\Delta^{\ddagger}G_{XY}$, and equilibrium, which is dependent on ΔG^0_{XY}, modulated by the mutual reorganization of solvation and/or structure as electrons flow from donor to acceptor along the reaction path [9].

The basic Marcus approach is not restricted to electron transfer and has been extended to other chemical reactions in solution that have collinear transition states [10–17] such as nucleophilic addition, methyl group transfer, proton transfer, and hydrogen atom transfer (proton-coupled electron transfer). Figure 2 illustrates the application of Marcus theory to the simple methyl group transfer reaction illustrated in Equation 4 [10–12]:

$$X^- + H_3C—Y \rightleftharpoons [^{\delta^-}X \cdots CH_3 \cdots Y^{\delta^-}]^{\ddagger} \rightleftharpoons X—CH_3 + Y^- \tag{4}$$
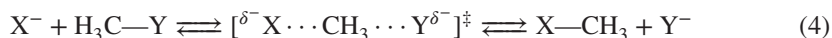
The Principle of Microscopic Reversibility dictates that the lowest Gibbs energy pathway from reagent to product is also necessarily the lowest Gibbs energy pathway in the reverse direction. Consequently, the Gibbs energy profile must be symmetrical for the identity reaction between CH_3–Y and Y^- (Equation 5), for which $\Delta G^0_{YY} \equiv 0$:

$$Y^- + H_3C—Y \rightleftharpoons [^{\delta^-}Y \cdots CH_3 \cdots Y^{\delta^-}]^{\ddagger} \rightleftharpoons Y—CH_3 + Y^- \tag{5}$$
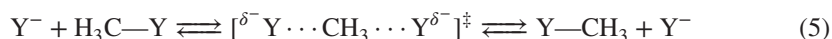
As shown in Figure 2a, the transition state must therefore occur at $P_{Rxn} = 0.50$, with a structure halfway between that of the reagents and products. From the crossing point of the two parabolas in Figure 2a, the Gibbs energy of activation of the reaction ($\Delta^{\ddagger}G_{YY}$) is found to be directly proportional to λ_{YY} (referred to as the solvent reorganization energy for electron transfer but here related to stretching of the CH_3–Y bond as well) and is given by Equation 6:

$$\Delta^{\ddagger}G_{YY} = \frac{\lambda_{YY}}{4} \tag{6}$$

For the analogous self-exchange reaction of H_3C–X with X^- (Equation 7):

$$X^- + H_3C—X \rightleftharpoons [^{\delta^-}X \cdots CH_3 \cdots X^{\delta^-}]^{\ddagger} \rightleftharpoons X—CH_3 + X^- \tag{7}$$
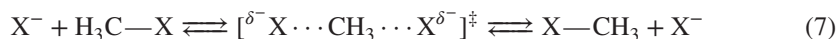
the equivalent expression is $\Delta^{\ddagger}G_{XX} = \lambda_{XX}/4$.
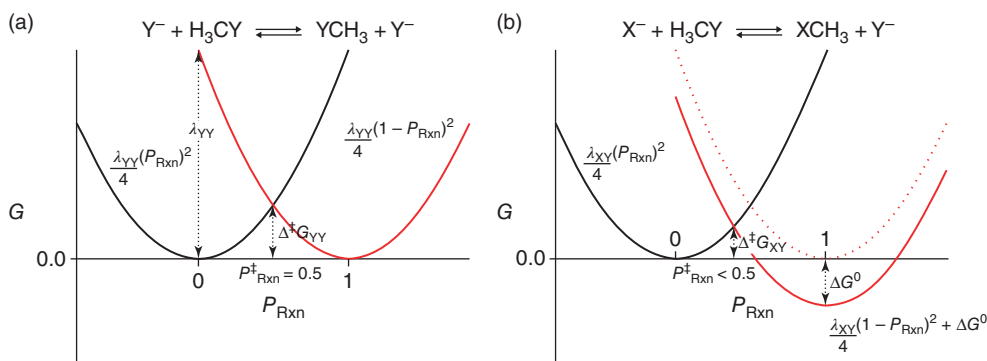
Figure 2   Marcus theory representation of the Gibbs energy profile as a function of the progress of the reaction in terms of parabolas: (a) for an isoergonic self-exchange reaction and (b) for an exergonic reaction.

TABLE 1   Application of Marcus Theory to Selected Methyl Transfer Reactions (Equation 4) in Water at 25 °C

X^-	CH_3–Y	$\Delta^{\ddagger}G_{XX}{}^a$	$\Delta^{\ddagger}G_{YY}{}^a$	$\Delta G^0_{XY}{}^a$	$P^{\ddagger}_{Rxn}{}^b$	$\Delta^{\ddagger}G_{XY}$ (Marcus)^c	$\Delta^{\ddagger}G_{XY}$ (Expt.)^a
Br^-	CH_3Br	99	99	0	0.50	99	99
CN^-	CH_3Br	213	99	−156	0.37	88	90
Br^-	$CH_3O_3S\phi$	99	149	−63	0.44	95	96
I^-	CH_3I	97	97	0	0.50	97	97
I^-	CH_3Br	97	99	−9	0.49	94	91
I^-	CH_3Cl	97	111	−3	0.50	103	100
I^-	CH_3F	97	133	−5	0.49	113	117

Gibbs energies in kilojoule per mole.
^{a}Reference [12].
^{b}Equation 9.
^{c}Equation 10.

Because the value of λ_{XY} for the cross reaction in Equation 4 should be intermediate between λ_{YY} and λ_{XX}, the reorganization energy for the cross reaction can be approximated as the average of the values of the two identity reactions, as shown in Equation 8:

$$\lambda_{XY} = \frac{(\lambda_{YY} + \lambda_{XX})}{2} \tag{8}$$

Employing this value of λ_{XY}, the reagent and product parabolas are initially constructed isoergonically; the product parabola is then shifted on the y-axis by the standard Gibbs energy difference (ΔG^0_{XY}) between the reagents and products, as indicated in Figure 2b. The new intersection point of the two parabolas whose equations are shown in Figure 2b provides the following expressions (Equations 9 and 10, respectively) for the progress of the reaction at the transition state ($P^{\ddagger}_{Rxn}$) and for the Gibbs energy of activation of the cross reaction ($\Delta^{\ddagger}G_{XY}$):

$$P^{\ddagger}_{Rxn} = 0.5 + \frac{\Delta G^0_{XY}}{2\lambda_{XY}} \tag{9}$$

$$\Delta^{\ddagger}G_{XY} = \frac{\lambda_{XY}}{4}\left(1 + \frac{\Delta G^0_{XY}}{\lambda_{XY}}\right)^2 \tag{10}$$

Equation 9 is consistent with the Hammond Postulate, that is, the progress of reaction at the transition state is <0.5 (more reagent-like) for an exergonic reaction and >0.5 (more product-like) for an endergonic reaction [5, 18, 19]. The data in Table 1 show the application of Equations 9 and 10 to some selected methyl transfer reactions (Equation 4) in water at 25 °C, emphasizing the expected trend in $P^{\ddagger}_{Rxn}$ with ΔG^0_{XY} and the generally good agreement between the calculated and experimental values of $\Delta^{\ddagger}G_{XY}$.

3.1   A Simple QSRR Based on Marcus Theory

Substituting Equations 6 and 8 into Equation 10 and rearranging, one obtains the following expression (Equation 11) for the methyl transfer reaction shown in Equation 4:

$$\Delta^{\ddagger}G_{XY} = \frac{\Delta^{\ddagger}G_{XX} + \Delta^{\ddagger}G_{YY} + \Delta G^0_{XY}}{2} + \frac{(\Delta G^0_{XY})^2}{4\lambda_{XY}} \tag{11}$$

For the reaction of iodide ion ($X = I^-$) with the four methyl halides (Table 1, last four entries), the approximation that $4\lambda_{XY} = \frac{\Delta^{\ddagger}G_{XX} + \Delta^{\ddagger}G_{YY}}{2} \gg (\Delta G^0_{XY})^2$ is valid and one would expect to observe an LFER of the form of Equation 12 [20]:

$$\Delta^{\ddagger}G_{XY} \approx \frac{\Delta^{\ddagger}G_{XX} + \Delta^{\ddagger}G_{YY} + \Delta G^0_{XY}}{2} = 48.5 + 0.50(\Delta^{\ddagger}G_{YY} + \Delta G^0_{XY}) \qquad (12)$$

Indeed, as shown in Equation 13, regression of the experimental values of $\Delta^{\ddagger}G_{XY}$ versus the sum of $(\Delta^{\ddagger}G_{YY} + \Delta G^0_{XY})$ does result in a rather good linear relationship (coefficient of determination, $R^2 = 0.97$; F-test $= 66$), despite the limited number of data points:

$$\Delta^{\ddagger}G_{XY} = 31.2 + 0.66(\Delta^{\ddagger}G_{YY} + \Delta G^0_{XY}) \qquad (13)$$

Comparison of the empirical relationship in Equation 13 with the relationship of Equation 12 derived from the theory indicates several general features that one can expect to encounter in QSRRs. Although $\Delta G^{\neq}_{YY}$ is the dominant factor [20], neither $\Delta G^{\neq}_{YY}$ nor ΔG^0_{XY} alone, but their sum, would be a completely adequate descriptor. With only four data points, the intercept and slope of the empirical correlation are different from the values expected from theory. Despite being intrinsically limited by the requirement that the nucleophile be the iodide ion, the empirical correlation does have some predictive utility for the reactions of other similar compounds with the iodide ion. Thus, for the reaction of the substrate CH_3ONO_2 ($\Delta^{\ddagger}G_{YY} = 111\,kJ\,mol^{-1}$; $\Delta G^0_{XY} = 0\,kJ\,mol^{-1}$), which was not used to develop the correlation, Equation 13 predicts a value of $\Delta^{\ddagger}G_{XY} = 104\,kJ\,mol^{-1}$ as compared to the experimental value of $103\,kJ\,mol^{-1}$ [12].

The following sections provide an overview of descriptors that have been successfully employed to develop chemically meaningful QSRR that relate the effects of substituents, nucleophilicity, and solvent on reactions and equilibria to molecular structure.

4   HAMMETT EQUATION AND SUBSTITUENT PARAMETERS

4.1   The Hammett Equation

In 1935, Hammett [21] and Burkhardt [22] independently reported the existence of linear log–log relationships between the rate constants for several different reactions of *m*- and *p*-substituted benzene derivatives and the corresponding equilibrium constants for a reference process such as the ionization of the corresponding *m*- and *p*-substituted benzoic acids. Although Burkhardt's brief report [22] reveals a clear perception of the implications of such correlations, it was Hammett who developed the first set of substituent constants and transformed quantitative correlations of substituent effects via such LFERs into an important mechanistic tool in organic chemistry [1, 23–27].

One of the early examples of an LFER was the correlation between the logarithms of the ionization constants of phenylacetic acids (X–PhAA) and the logarithms of the ionization constants of the corresponding substituted benzoic acids (X–BzA) at 25 °C in aqueous solution [22, 26] of the form shown in Equation 14:

$$\log K_{a(\text{X–PhAA})} = \rho \log K_{a(\text{X–BzA})} + \text{const} \qquad (14)$$

By referencing the substituent effects to those of the unsubstituted analogs (X = H), Equation 14 can be rewritten as Equation 15:

$$\log K_{a(X—PhAA)} - \log K_{a(H—PhAA)} = \rho[\log K_{a(X—BzA)} - \log K_{a(H—BzA)}] = \rho\sigma \quad (15)$$

where the substituent constants σ were defined by Hammett [27] as shown in Equation 16:

$$\sigma = \log K_{a(X—BzA)} - \log K_{a(H—BzA)} = pK_{a(H—BzA)} - pK_{a(X—BzA)} = \frac{-\delta\Delta G^0}{2.303RT} \quad (16)$$

and denoted as either σ_m or σ_p for substituents at the *m*- or *p*-positions of benzoic acid, respectively. The values of σ are directly related to differences in the Gibbs energy of ionization, $\delta\Delta G^0$, due to the presence of the substituents, that is, to the net difference in stabilization of the ionized and unionized forms of benzoic acids by the substituent compared to that of benzoic acid itself (Figure 3).

For the ionization of benzoic acids, the magnitude of the stabilization is relatively small, a change of ± 1 in σ corresponding to an increment of only $\pm 5.8 \, \text{kJ mol}^{-1}$ ($\pm 1.4 \, \text{kcal mol}^{-1}$) in $\delta\Delta G^0$. The enthalpies of dissociation of benzoic acids in water are rather small and positive, while the corresponding entropy changes are relatively large and negative [28] and hence determine the values of σ. Thus, for the ionization of benzoic acid itself, $\Delta H^0 = 0.5 \, \text{kJ mol}^{-1}$ ($0.11 \, \text{kcal mol}^{-1}$) as compared to $-T\Delta S^0 = 23.5 \, \text{kJ mol}^{-1}$ ($5.6 \, \text{kcal mol}^{-1}$) [28]. Electron-donating substituents increase the negative charge density on the carboxylate group, resulting in a greater demand for solvation, while electron-withdrawing substituents decrease the charge density, with a corresponding decrease in the entropy of solvation of the benzoate anion. Net electron-withdrawing substituents that facilitate the ionization of benzoic acid (stabilize the benzoate anion) thus have positive values of σ, whereas net electron-donating substituents that destabilize the benzoate ion relative to benzoic acid have negative values of σ (Figure 3). The magnitude and sign of ρ thus provide a quantitative measure of the relative sensitivity of the reaction
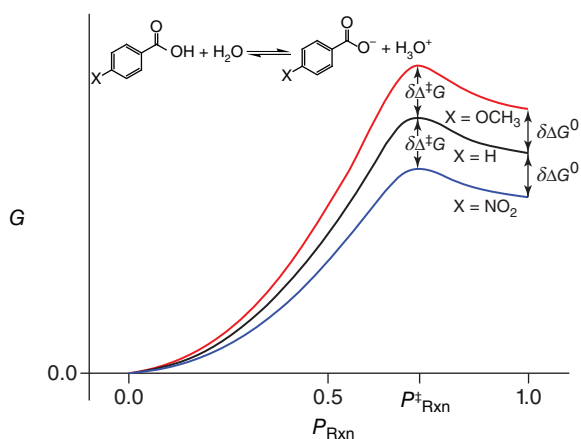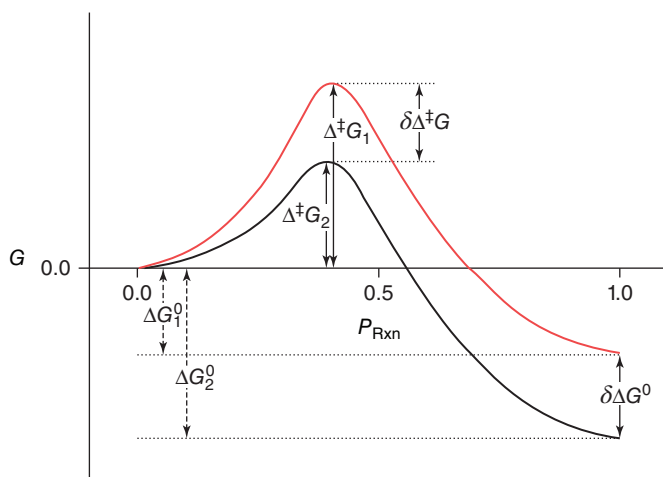

Figure 3 Schematic representation of the differential effect of electron-donating (X = OCH$_3$) and electron-withdrawing (X = NO$_2$) para-substituents on the Gibbs energy profile for the acid–base equilibrium of benzoic acids.

or equilibrium of interest to substituents as compared to the ionization of benzoic acids. As an example, comparison of the slopes of plots of log K_a for the ionization of phenylacetic acids ($\rho = 0.49$) and 3-phenylpropionic acids ($\rho = 0.21$) with that of benzoic acids ($\rho = 1.00$ by definition) [26] indicates that the insertion of methylene groups between the aromatic ring and the carboxylate group decreases the transmission of the substituent effect by a factor of roughly two for each CH_2 group.

The choice of benzoic acids as the reference series for the determination of σ values was both logical and perhaps serendipitous. From a practical standpoint, benzoic acids are relatively simple to prepare and the ionization constants can be easily determined with the requisite precision at 25 °C. At the same time, the equilibrium constant for ionization is a function of the ratio of rate constants for deprotonation (k_{deprot}) of the carboxylic acid and protonation (k_{prot}) of the carboxylate anion, that is, $K_a = k_{deprot}/k_{prot}$. The protonation of oxyanions such as carboxylates and phenoxides is extremely fast [29, 30] and k_{prot} is insensitive to the presence of substituents ($k_{protX} \sim k_{protH}$). This reflects the fact that the protonation is effectively diffusion controlled (unit probability of reaction per encounter) rather than the "Reactivity–Selectivity Principle" (increased reactivity implies decreased selectivity), an otherwise attractive concept that has been largely discredited as a general principle [31]. Since the activation barrier for protonation is small, the Hammond Postulate requires that the transition state for protonation be very similar to that of the anion. As a consequence (Figures 3 and 4), the Hammett σ values are not only related to the ratio of equilibrium constants for ionization but can also be conveniently related to the log of the ratio of rate constants for deprotonation of the benzoic acids (Equation 17):

$$\sigma = \log\left(\frac{K_{a(X-BzA)}}{K_{a(H-BzA)}}\right) \approx \log\left(\frac{k_{deprotX}}{k_{deprotH}}\right) = \delta\Delta^{\ddagger}G_{deprot} \tag{17}$$

and hence to the differential effect of substituents on the Gibbs energy of activation for deprotonation of the reference benzoic acids.

Figure 4 Schematic representation of the influence of a substituent on the Gibbs energy of activation for an exergonic reaction.

When applied to substituent effects on reaction rates, the Hammett equation correlates the logarithm of the ratio of the rate constants for reactants with (k_X) and without (k_H) substituents to σ, as illustrated in Equation 18:

$$\log\left(\frac{k_X}{k_H}\right) = \rho\sigma \tag{18}$$

In interpreting linear Hammett correlations, one is implicitly assuming that [1, 2] (i) electronic and steric effects are separable and additive; (ii) the σ values adequately encode only these electronic effects; (iii) neither the rate-determining step nor the mechanism of the reaction change as the substituent is varied; and (iv) the progress of the reaction at the transition state does not change significantly with substituent. For linear plots, the slope ρ provides a measure of the magnitude and direction of the development of charge in the transition state relative to that involved in the ionization of benzoic acids. A positive value of ρ can be associated with net development of negative charge at the reaction center in the transition state of the rate-limiting step and a negative value of ρ with net development of positive charge at the reaction center in the transition state of the rate-limiting step (Figure 5). Experimental values of ρ for a series of typical organic reactions are indicated in Scheme 1.

Nonlinear Hammett plots that curve at the extremities can reveal substituent-dependent variations in the electron demand at the reaction center. In these cases, free energy correlations employing Hammett substituent constants that take into account effects such as direct resonance between the substituent and the reaction center (see below, through resonance interactions) will often linearize the free energy relationship [32]. In contrast, more abrupt changes in the slope, resulting in upwardly concave Hammett plots, are usually taken to be indicative of a change in mechanism, whereas downward concave plots are usually considered to be an indication of a change in the rate-determining step of the reaction rather
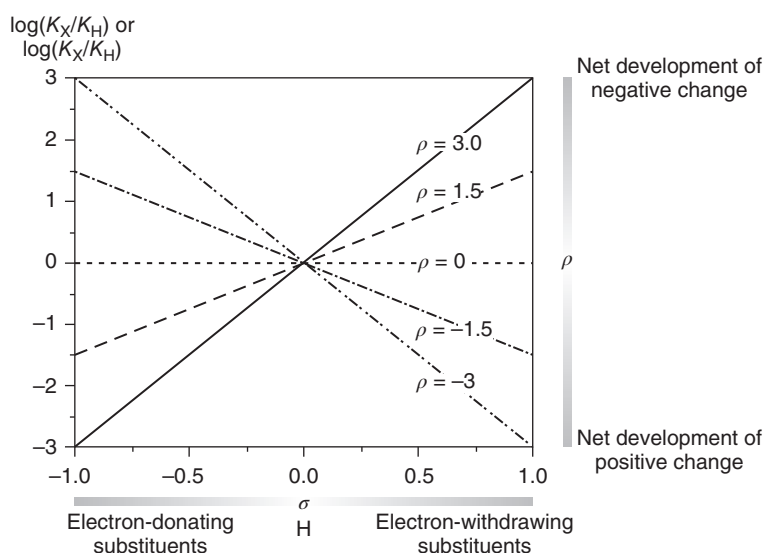
Figure 5   Typical range of values of ρ for linear Hammett plots and the corresponding interpretation in terms of charge development at the reaction center.
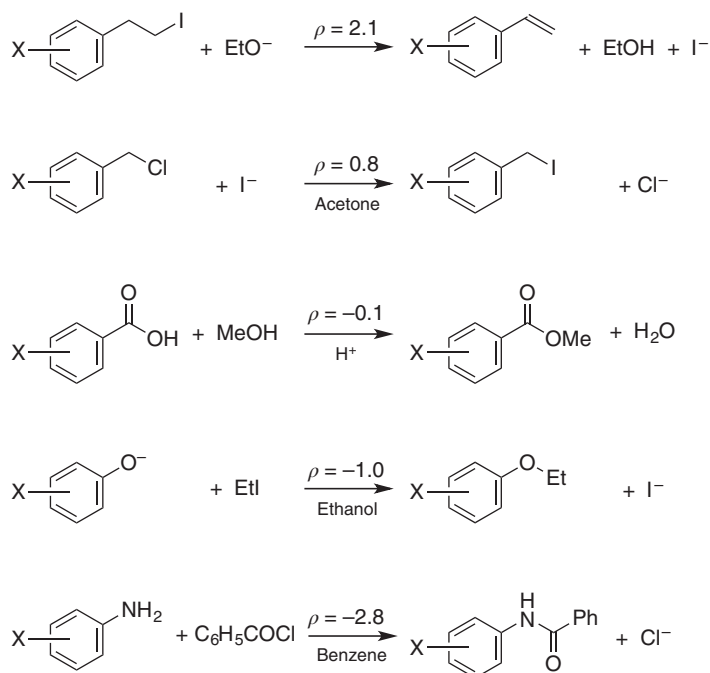
Scheme 1   Examples of chemical reactions with values of ρ ranging from positive to negative [1, 2].

than the mechanism. A classical example of a change in the rate-determining step is the uncatalyzed reaction between substituted benzaldehydes and a primary alkyl amine to form the corresponding imine (Figure 6). With electron-withdrawing substituents ($\sigma > 0$), formation of the tetrahedral intermediate (step 1 in Figure 6) is rate limiting, whereas for electron-donating substituents ($\sigma < 0$), it is the dehydration of the tetrahedral intermediate to form the imine (step 2 in Figure 6) that becomes rate limiting [33].

The comparison of the hydrolysis of the methyl and ethyl benzoates in concentrated sulfuric acid at 45 °C [34] provides a striking contrast between the linear Hammett plot for the methyl benzoates and the upwardly concave plot for the ethyl benzoates (Figure 7). With electron-donating and weakly electron-withdrawing substituents, both the methyl and ethyl esters hydrolyze via carbonyl-carbon/alcohol-oxygen cleavage (steps 1 and 2 in Figure 7). With strongly electron-withdrawing substituents, the mechanism remains the same for the methyl esters; however, for the ethyl esters, the dominant mechanism becomes oxygen–carbon cleavage, producing the ethyl carbocation (step 3 in Figure 7).

Although Hammett σ values for ortho substituents, σ_{ortho}, can be readily determined from the ionization constants of the corresponding benzoic acids and have been tabulated [35, 36], their utility for understanding electronic effects on reactivity and equilibria is limited by their proximity to the reaction center. Thus, depending on the substituent, the σ_{ortho} values themselves may encode a mixture of electronic and steric effects [37, 38]. When the reaction of interest has important steric requirements, σ_{ortho} values alone can be expected to be inadequate without an additional steric hindrance parameter. Nonetheless, attempts to include such a parameter have not been particularly fruitful, suggesting that the electronic and steric effects of ortho substituents may not be additive [39, 40]. Not surprisingly, ortho
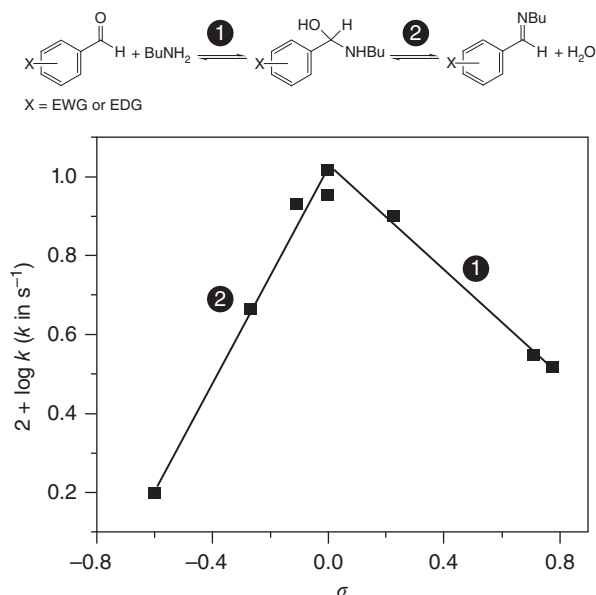
Figure 6 Hammett plot for the uncatalyzed reaction between substituted benzaldehydes and *n*-butyl amine to form the corresponding imine [33], indicating the change in rate-determining step from formation of the tetrahedral intermediate (step 1) to dehydration of the tetrahedral intermediate to form the imine (step 2).

substituents are usually excluded from Hammett-type correlations. Other potential problems arising from close proximity to the reaction center include specific interactions that are not adequately represented by either σ_{ortho} or by an additional steric parameter such as steric hindrance of the resonance of an adjacent substituent or hydrogen bonding between an ortho substituent and a group that is directly involved in the reaction.

Another limitation for the practical application of the Hammett equation concerns the proper choice of σ values when multiple substituents are present. Although it is often assumed that σ values are additive [26, 41], this is not necessarily the case, especially when the substituents are in close proximity, where additional steric effects can come into play. This problem has been considered in more detail by Kalfus *et al*. [42], and there are limited compilations of effective σ values for some multiple substituent combinations. An analogous problem occurs in the choice of effective σ values for polycyclic aromatic systems such as naphthalenes [43, 44] or flavyliumcations [45] with multiple substituents on different rings.

4.2   Temperature Dependence of the Hammett ρ Value

A still poorly understood aspect of Hammett plots is the fundamental significance of the variation of the slope ρ with temperature, which was explicitly pointed out by Burkhardt in 1935 [22]. The Hammett equation (Equation 18) depends on the difference in the Gibbs energy of activation with and without the substituent, $\delta\Delta^{\ddagger}G_{\text{X–H}}$, as shown in Equation 19:

$$\delta\Delta^{\ddagger}G_{\text{X—H}} = \delta\Delta^{\ddagger}H_{\text{X—H}} - T\delta\Delta^{\ddagger}S_{\text{X—H}} = -2.303RT\rho\sigma \tag{19}$$
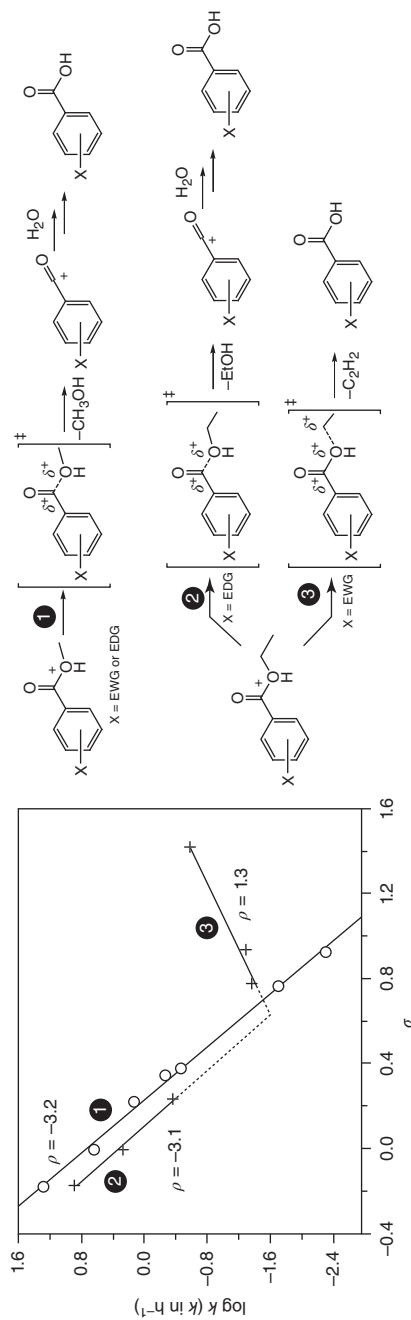
Figure 7 Hammett plot and hydrolysis mechanism of methyl and ethyl benzoates in concentrated sulfuric acid (45 °C) [34], showing the linear plot for acyl carbon–alcohol oxygen cleavage for the methyl esters (step 1) and the change in mechanism from acyl carbon–alcohol oxygen (step 2) to carbon–oxygen cleavage to form the ethyl carbocation (step 3).

Experimentally, it is often found that enthalpies of activation, $\delta\Delta^{\ddagger}H_{X-H}$, correlate with the corresponding entropies of activation, $\delta\Delta^{\ddagger}S_{X-H}$ via a relationship of the form of Equation 20:

$$\delta\Delta^{\ddagger}H_{X-H} = T_{comp}\delta\Delta^{\ddagger}S_{X-H} \tag{20}$$

Indeed, this enthalpy–entropy compensation may very well be a requisite condition for the existence of an LFER [6, 7]. The proportionality constant T_{comp} in Equation 20 is the compensation temperature [46, 47]. When the Arrhenius plots for the kinetics of the reaction of interest converge to a common point for all substituents, the corresponding temperature β_{iso} is known as the isokinetic temperature (or in the case of an equilibrium process, as the isoequilibrium temperature) [8, 46–55]. Combining these two equations with $T_{comp} = \beta_{iso}$, one obtains Equation 21:

$$\delta\Delta^{\ddagger}G_{X-H} = \left(1 - \frac{T}{\beta_{iso}}\right)\delta\Delta^{\ddagger}H_{X-H} \tag{21}$$

which predicts that the substituent effect on the Gibbs energy of activation, and hence on the reaction rate, should disappear at the isokinetic temperature, that is, ρ should be zero when $T = \beta_{iso}$ and change sign at temperatures above the isokinetic temperature. Thus, when a substituent effect is absent, care must be taken to ensure that the experimental temperature is not close to T_{comp} or β_{iso}, although in most cases isokinetic temperatures are in the range of several hundreds of degree celsius [47, 54]. Since the Hammett slope ρ is proportional to $-\delta\Delta^{\ddagger}G_{X-H}/RT$ (Equation 19), ρ is expected to be inversely proportional to the temperature [56], as indicated in Equation 22:

$$\rho(T) = \rho_{\infty}\left(1 - \frac{\beta_{iso}}{T}\right) \tag{22}$$

where ρ_{∞} is the corresponding limiting value of the slope at $T \rightarrow \infty$. This relationship has been shown to hold for the specific case in which $\delta\Delta^{\ddagger}H_{X-H}$ and $\delta\Delta^{\ddagger}S_{X-H}$ are temperature-independent (constant differential heat capacity) [51]. However, when these vary with temperature, the temperature dependence of ρ is more complex and β_{iso} itself becomes temperature dependent [51].

5   THE PANTHEON OF HAMMETT-RELATED SUBSTITUENT PARAMETERS

Substituent effects on reactions and equilibria depend on a variety of factors, including inductive or electric field effects exerted on the reaction center by the substituent, resonance interactions, temperature, polarizability, solvent (especially for ionic substituents, where ionic strength can be a complicating factor), and steric interactions. The widespread application of the Hammett equation to correlate reactivity patterns soon identified some of these limitations of the original Hammett σ_m and σ_p constants based on the ionization of benzoic acids. In 1968, Swain and Lupton [57] noted that at least 20 different scales of Hammett σ constant had been proposed in the literature. The most important of these alternative σ constant scales are considered in the following sections.

5.1   Through-Resonance Interactions

Hammett himself noted deviations for strongly electron-withdrawing *p*-substituents in Hammett plots for the ionization of phenols, attributable to direct conjugation of the substituent with the reaction center (through-resonance), as illustrated in Scheme 2 for the *p*-nitrophenoxide ion.
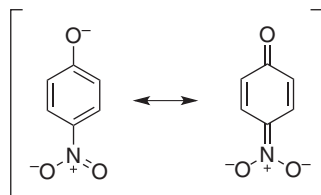
Scheme 2   Resonance structures of the *p*-nitrophenoxide ion.

This led to the development of the σ_p^- scale, originally based on the ionization of phenols but later on anilines [35, 36]. The values of σ_p^- are much larger than σ_p for highly electron-withdrawing para-substituents, but comparable to those of σ_p for electron-donating substituents (Table 2). The Hammett σ_p values for strongly electron-donating substituents such as –NH$_2$, –OH, and –OR are themselves subject to a through-resonance interaction with the carboxylate group, as indicated in Scheme 3 for the *p*-methoxybenzoate ion.
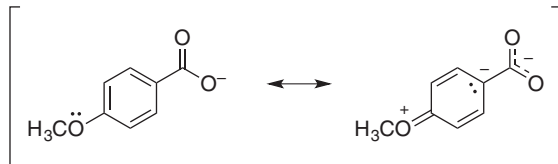
Scheme 3   Resonance structures of the *p*-methoxybenzoate ion [2].

For these few substituents, the values of σ_p^0 (Table 2) based on reference systems such as the ionization of phenylacetic acids, where an intervening methylene group insulates the carboxylate ion from the substituent, can be used to correct σ_p for through-resonance interactions [36]. The through-resonance interaction of these substituents with centers of positive charge is even greater than that with benzoate ions, indicating the need for a corresponding set of σ_p^+ values (Table 2). For these substituents, Okamoto and Brown [59] developed such a set of values based on the solvolysis of cumyl chlorides in 90% aqueous acetone (Scheme 4) rather than on an equilibrium constant. The σ_p^+ constants were then defined via Equation 23:

$$\sigma_p^+ = \frac{\log(k_X/k_H)}{4.54} \tag{23}$$

where the numerical factor in the denominator scales the values so that they are consistent with the Hammett constants for meta-substituents (making $\sigma_m^+ = \sigma_m$). A recent interesting approach for the determination of σ^+ values in nonpolar solvents is via the UV–visible absorption maximum of substituted nitrobenzenes [60].

TABLE 2   Values of Hammett Substituent Constants for Para- (σ_p) and Meta- (σ_m) Substituents, Para-substituent Constants for Through-Resonance with Positive (σ_p^+) Negative (σ_p-) Reaction Centers, σ_p Values Corrected for through-resonance (σ_p^o), Inductive Substituent Constants (σ_I), Swain–Lupton Field (F) and Resonance (R) Parameters, Taft Polar (σ^*) and Steric (E_S) Constants, and Para-substituent Constants for Stabilization of Free Radicals (E_R) for a Selection of Commonly Employed Substituents

	Hammett		Through resonance			Inductive	Swain–Lupton		Taft polar/steric		Radicals
Substituent	σ_p	σ_m	σ_p^+	σ_p-	σ_p^o	σ_I	F	R	σ^*	E_S	$\sigma_{JJ}^\bullet$
–N(CH$_3$)$_2$	−0.83	−0.16	−1.7	−0.12	−0.48	0.06	0.15	−0.98	—	—	1.00
–NH$_2$	−0.66	−0.16	−1.30	−0.15	−0.36	0.12	0.08	−0.74	0.62	−0.61	—
–OH	−0.37	0.12	−0.92	−0.37	−0.16	0.29	0.33	−0.70	1.37	−0.55	—
–OCH$_3$	−0.27	0.12	−0.78	−0.26	−0.15	0.27	0.29	−0.56	1.77	−0.55	0.23
–C(CH$_3$)$_3$	−0.20	−0.10	−0.26	−0.13	−0.17	−0.07	−0.02	−0.18	−0.07	−2.78	0.26
–CH$_3$	−0.17	−0.07	−0.31	−0.17	−0.12	−0.04	0.01	−0.18	0	−1.24	0.15
–C$_6$H$_5$	−0.01	0.06	−0.18	0.02	0.01	0.12	0.12	−0.13	0.60	—	0.47
–H	0	0	0	0	0	0	0	0	0.49	0	0
–SCH$_3$	0.00	0.15	−0.60	0.06	−0.02	0.25	0.23	−0.23	1.56	−1.07	0.62
–F	0.06	0.34	−0.07	−0.03	0.21	0.52	0.45	−0.39	3.19	−0.55	−0.02
–I	0.18	0.35	0.14	0.27	0.31	0.39	0.42	−0.24	2.22	−1.62	—
–Cl	0.23	0.37	0.11	0.19	0.28	0.47	0.42	−0.19	2.94	−0.97	0.22
–Br	0.23	0.39	0.15	0.25	0.28	0.44	0.45	−0.22	2.80	−1.16	0.23
–CO$_2$CH$_3$	0.45	0.36	0.49	0.75	0.46	0.32	0.34	0.11	2.00	—	0.33
–COCH$_3$	0.50	0.38	—	0.84	0.50	0.30	0.33	0.17	1.65	—	0.54
–CF$_3$	0.54	0.43	0.61	0.65	0.54	0.40	0.38	0.16	2.85	−2.40	−0.01
–NH$_3^+$	0.60	0.86	—	−0.56	—	0.61	0.92	−0.32	3.61	—	—
–CN	0.66	0.56	0.66	1.00	0.68	0.53	0.51	0.15	3.64	−0.51	0.42
–NO$_2$	0.78	0.71	0.79	1.27	0.82	0.64	0.65	0.13	4.66	−2.52	0.36
–N(CH$_3$)$_3^+$	0.82	0.88	0.41	0.77	0.37	0.93	0.86	−0.04	4.16	—	—

Data from the compilations of Hansch *et al.* [35, 36], except for $\sigma_{JJ}^\bullet$ from Jiang [58].
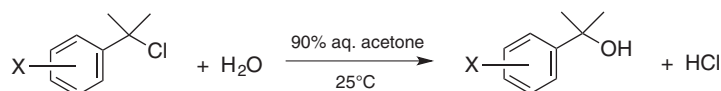
Scheme 4   Solvolysis of cumyl chlorides, used to develop the σ^+ scale [59].

The σ^+ constants have been shown to be more appropriate than the original Hammett σ values for reactions that involve significant positive charge stabilization by substituents, as indicated by classical examples such as electrophilic aromatic substitution [61] and the acid-catalyzed hydration of 1,1-disubstituted ethylenes [62]. The difference between correlations based on conventional Hammett σ constants and σ^+ values is illustrated in Figure 8 for the acetolysis of substituted brosylates.

The relative importance of through-resonance can be made explicit in Hammett plots by employing Equation 24, known as the Yukawa–Tsuno equation [64, 65]:

$$\log\left(\frac{k_X}{k_H}\right) = \rho[\sigma + r(\sigma^+ - \sigma)] \tag{24}$$
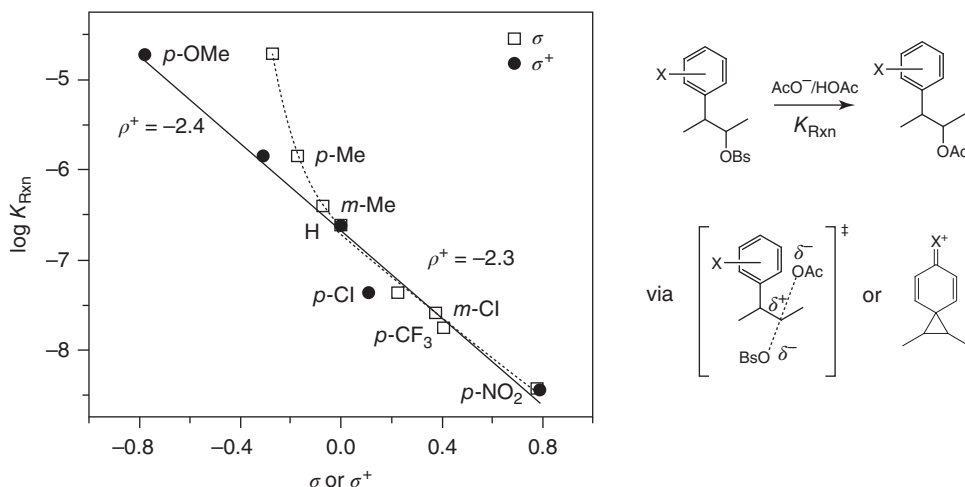
Figure 8   Comparison of the curved Hammett plot for the acetolysis of substituted brosylates based on conventional Hammett σ_m and σ_p values with the linear free energy relationship based on the corresponding σ^+ values [63], indicating the importance of through-resonance interactions for the electron-donating substituents.

or its counterpart with σ^- in place of σ^+ [60, 66].

5.2   Separation of Resonance and Field-Inductive Contributions

As implied in the previous section, the Hammett constants contain contributions from both resonance and inductive or field effects, with resonance interactions being more important in σ_p than in σ_m. Factoring the Hammett σ values into resonance (σ_R) and inductive (σ_I) contributions leads to Equation 25:

$$\sigma = \sigma_R + \sigma_I \tag{25}$$

which requires reliable, properly scaled values of σ_I in order to determine σ_R (or vice versa).

For nonaromatic systems, Taft [67, 68] determined a set of polar or inductive substituent constants designated as σ^* values (Equation 26) derived from the base- (B) and acid- (A) catalyzed hydrolysis of substituted methyl or ethyl acetates ($X–CH_2CO_2–R$) with X=H as the reference rate constant:

$$\sigma^* = \frac{\left[\log\left(\frac{k_X}{k_H}\right)_B - \log\left(\frac{k_X}{k_H}\right)_A\right]}{2.48} \tag{26}$$

The numerical factor in the denominator serves to scale the σ^* values to the Hammett σ values. The term for the acid-catalyzed hydrolysis, which has a ρ value close to zero, was subtracted from that of the base catalyzed hydrolysis in order to correct for the steric effects on the latter and was denoted as E_S (Equation 27):

$$E_S = \frac{\log\left(\frac{k_X}{k_H}\right)_A}{2.48} \tag{27}$$

The relative importance of polar and steric effects for nonaromatic systems can then be gauged by the Hammett-type correlation shown in Equation 28:

$$\log\left(\frac{k_X}{k_H}\right) = \rho^*\sigma^* + \delta_S E_S \tag{28}$$

where ρ^* and δ_S are the factors that quantify the sensitivity of the reaction rate to polar and steric effects.

Other sets of compounds that have been employed to develop inductive parameters for substituents include 4-substituted-bicyclo[2.2.2]octane carboxylic acids (known as σ' values) [69], 4-substituted quinuclidines (σ_{IQ} values) [70, 71], and substituted acetic acids (Charlton σ_I values) [72]. Intercorrelation and scaling of the former two values led to the following relationship (Equation 29) for σ_I [35]:

$$\sigma_I = 1.297\sigma_m - 0.385\sigma_p + 0.033 \tag{29}$$

Attempts to factor the inductive effect further into through-bond and through-space (or perhaps more precisely, through-solvent) components have not proved to be particularly fruitful, perhaps because of a predominance of the through-space component [39, 73].

5.3   The Swain–Lupton Equation

An alternative approach for separating field and resonance effects was proposed by Swain and Lupton [57, 74] in the form of substituent-position-independent field (F) and resonance (R) parameters that were employed in the modified Hammett equation depicted in Equation 30:

$$\log\left(\frac{k_X}{k_H}\right) = \rho\sigma = fF + rR \tag{30}$$

where the slopes f and r indicate the sensitivity of the reaction center to the field and resonance effects, respectively. Using the assumptions that $F = \sigma' \approx \sigma_I$ and $R = 0$ for the substituent trimethylammonium, Swain and Lupton developed the following relationship (Equation 31) between σ_p, F, and R:

$$\sigma_p = 0.92F + R \tag{31}$$

from which the values of F and R could be calculated. These two parameters have the additional advantage that they can be related to all the other major types of σ values introduced earlier, obviating the need to decide which set of σ values is the most appropriate for analyzing the reactivity data at hand.

5.4   Quantum Chemical Calculation of Hammett Substituent Constants

The increasing availability of more accurate quantum chemical methods has led to an upsurge in attempts to relate quantum chemical molecular parameters to the various types of substituent constants, with varying degrees of success [39, 75–77]. Among the most successful predictors of Hammett σ values is the difference between the core electron binding energies (ΔCEBE) of the corresponding o-, m-, or p-carbon of a substituted benzene compared to that of the unsubstituted compound (e.g., C_6H_5–X relative to C_6H_6), which can be

calculated by density functional theory [78, 79]. Takahata [80, 81] has separated resonance and inductive effects by showing that the ΔCEBE of substituted cyclohexanes (C_6H_{11}–X) correlates with σ_I and the difference between the ΔCEBE of substituted benzenes and the corresponding substituted cyclohexane correlates with σ_R.

5.5   Hammett Constants for Free Radical Substituents and for the Stabilization of Free Radical Reaction Centers

Theoretical calculations of the experimentally inaccessible pK_as of benzoic acids bearing free radical substituents have been employed to develop a set of Hammett σ constants for the stabilization of reaction centers by free radical substituents [82]. Substituent effects on aliphatic free radical centers have been correlated with Taft σ^* values [83, 84] and substituent effects on radical centers in resonance with aromatic rings have been tied to σ^+ [84, 85]. Although substituent constants for the stabilization of radical centers have been estimated, for example, from measurements of the rate constants for hydrogen atom abstraction from substituted cumenes (E_R values; Table 2) and from the EPR hyperfine coupling constants of benzylic hydrogens ($\sigma^\bullet$ values) [36], adequate scaling and separation of polar and radical effects is lacking. Jiang, however, has developed a consistent set of substituent constants for stabilization of para-substituted benzylic radical centers, the $\sigma_{JJ}^\bullet$ scale (Table 2), which explicitly factors out the polar effects of the substituent [58]. A consistent feature of all of these scales is that most substituents have a net stabilizing influence on the radical center relative to hydrogen.

5.6   Sources of Substituent Parameters

The most comprehensive collections of substituent parameters have been compiled by Hansch *et al.* [35, 36]. Values of Hammett substituent constants for para- (σ_p) and meta- (σ_m) substituents, para-substituent constants for through-resonance with positive (σ_p^+) and negative (σ_p^-) reaction centers, σ_p values corrected for through-resonance (σ_p^0), inductive substituent constants (σ_I), Swain–Lupton field (F) and resonance (R) parameters, Taft polar (σ^*) and steric (E_S) constants, and para-substituent constants for stabilization of free radicals (E_R) are presented in Table 2 for a selection of commonly employed substituents.

5.7   Hammett Correlations for Electronically Excited States

There have been scattered attempts to use Hammett correlations to interpret substituent effects on the reactivity of electronically excited states or to develop excited-state substituent constants [75, 86, 87]. However, the fact that the energy and orbital origin (n,π^* or π,π^*) of the lowest electronically excited state are inherently substituent-dependent undoubtedly dooms any attempt to develop a generally applicable set of excited-state substituent constants. On the other hand, a Hammett correlation was successfully used to rationalize the chemiexcitation mechanism for the formation of singlet excited states in the decomposition of electron-rich 1,2-dioxetanes [88].

5.8   Some Unique Applications of Hammett Correlations

Several recent examples of unique applications of Hammett sigma constants underscore their continuing importance for a deeper understanding of electronic effects on reactivity and equilibria. Hammett correlations have been extensively employed to investigate

arene–arene, cation–arene, and anion–arene noncovalent bonding interactions [89] and, more recently, the performance of asymmetric catalysts [90]. Hammett plots for substituted iodobenzenes, in combination with quantum chemical calculations, indicated that the rate-limiting step of the Sonogashira coupling reaction is the initial oxidative addition of palladium to the carbon-iodine bond [91]. A particularly notable application was the demonstration by Rao *et al.* [92] that the Hammett equation could be applied to correlate and predict substituent effects on the single-molecule chemistry of substituted thiophenols adsorbed onto a copper surface.

6   PARAMETERS FOR STERIC EFFECTS

The Taft steric parameter E_S was introduced in Equation 27. As originally defined, the reference substituent was the methyl group. In analyzing LFERs containing E_S as a descriptor, attention must be paid to the values employed since the values of E_S were rescaled so that the reference substituent was hydrogen, as indicated in Equation 32:

$$E_S \quad (\text{rescaled}) = E_S \quad (\text{original}) - 1.24 \tag{32}$$

A number of other steric or size-related parameters have been proposed [93], including Charton's minimum van der Waals radius parameter for symmetrical substituents (Equation 33):

$$v_X = r_{VX} - r_{VH} = r_{VX} - 1.20 \tag{33}$$

which exhibits a strong correlation with E_S, and Hancock's correction of E_S values for the number n_H of alpha hydrogens (Equation 34):

$$E_S^C = E_S + 0.306(n_H - 3) \tag{34}$$

an approach extended by Fujita to unsymmetrical substituents of the type $-CR^1R^2R^3$ using the weighted sum of the E_S^C values for the three R groups.

Verloop's STERIMOL [94] parameters are particularly attractive because they take into account the three-dimensional aspects of substituent shape and are derived directly from molecular structure. The three principal parameters of STERIMOL are the length (L), minimal width (B_1), and maximal width (B_5) of the substituent group. The STERIMOL program based on the original algorithm for calculating the Verloop parameters is still available in a FORTRAN version from the Quantum Chemistry Program Exchange(Program Number QCMP093) [95]. Despite their limitations, the STERIMOL parameters have been shown to be useful for analyzing reactions involving asymmetric catalysis [96].

6.1   Polarizability

The most readily calculable polarizability parameter is the Molar Refraction, MR, given by the relationship in Equation 35 [36]:

$$\text{MR} = \frac{(n^2 - 1)}{(n^2 + 1)} \frac{\text{MW}}{d} \tag{35}$$

where n is the refractive index and MW/d the molecular weight divided by the density of the substance of interest. In practice, the value of MR is often scaled by a factor of 1/10.

7   NUCLEOPHILICITY AND ELECTROPHILICITY SCALES

At some point along the reaction coordinate of the reaction between a nucleophile and an electrophile, there must be partial electron transfer from the nucleophile to an unfilled orbital of the electrophile. The details of this transfer will depend on the mixing between the highest occupied molecular orbital of the nucleophile ($HOMO_{Nu}$) and the lowest unoccupied molecular orbital of the electrophile ($LUMO_E$), which in turn will depend on the difference in energy between them. Clearly, any attempt to quantify nucleophilicity can be expected to be electrophile dependent. Therefore, the premise that one can construct absolute scales of nucleophilicity is fundamentally flawed [97]. Despite this intrinsic limitation, much progress has been made in developing quantitative treatments of nucleophilicity and electrophilicity.

Swain and Scott [98] used the expression in Equation 36:

$$\log\left(\frac{k}{k_0}\right) = sn \tag{36}$$

to correlate rate constants (under pseudo-first-order reactionconditions) for nucleophilic attack on a substrate relative to that (k_0) of nucleophilic attack on the same substrate in water at 25 °C. The nucleophilicity constants n were determined from the rate of reaction of methyl bromide, for which the slope s was defined to be unity. Like ρ in the Hammett equation, the value of s expresses the sensitivity of the electrophile to n relative to that of methyl bromide. On the other hand, the Ritchie equation [99, 100], shown in Equation 37:

$$\log\left(\frac{k}{k_0}\right) = N^+ \tag{37}$$

based on nucleophilic reactivity with tritylium ions, omitted the electrophile-dependent sensitivity term, tantamount to assuming that the nucleophilicity was independent of the nature of the electrophile, that is, that an absolute nucleophilicity scale could in fact exist.

The Mayr–Patz equation [97, 101, 102], shown in Equation 38:

$$\log(k) = s_N(N + E) \tag{38}$$

has been employed to correlate rate constants for reactions of a wide range of π-, n-, and σ-nucleophiles (alkenes, arenes, enol ethers, enamines amines, alcohols, inorganic anions, transition metal complexes, hydrides, etc.) with diarylmethylcationsor neutral quinine methides with rate constants spanning 30 orders of magnitude. The parameters N and E are the nucleophilicity and electrophilicity of the reaction partners and s_N is the nucleophilic-dependent slope, taken to be unity for the standard π-nucleophile 2-methyl-1-pentene. In this equation, both N and s_N are, in principle, solvent dependent, whereas E is considered to be a uniquely defined solvent-independent parameter for the electrophile. The value of E was set to zero for the reaction of 2-methyl-1-pentene (s_N = 1.0) with the cation ($4\text{-MeOC}_6\text{H}_4)_2\text{CH}^+$. The Mayr group maintains a database of nucleophilicities and electrophilicities [103].

Mayr and Ofial [97] pointed out that all of these previous treatments can be unified as limiting cases of the following general equation (Equation 39):

$$\log(k) = s_N s_E(N + E) \tag{39}$$

where s_E is an additional electrophile-dependent parameter. With $s_E = 1$ for carbocations, this equation reduces to the Mayr–Patz equation (Equation 38). For most of the n-nucleophiles examined by Swain and Scott, $s_N \approx 0.6$ and Equation 39 reduces to Equation 40:

$$\log(k) = 0.6s_E E + 0.6s_E N \qquad (40)$$

For a single electrophile (constant E), the inclusion of the reference reaction with water as nucleophile leads to a relationship formally equivalent to the Swain–Scott equation (Equation 36):

$$\log\left(\frac{k}{k_0}\right) = 0.6s_E N \qquad (41)$$

Finally, assuming that the reactivity is electrophile-independent ($s_E = 1.0$), Equation 41 is further simplified to Equation 42, which is identical in form to the Ritchie equation (Equation 37).

$$\log\left(\frac{k}{k_0}\right) = 0.6N \qquad (42)$$

A Frontier molecular orbital interpretation of the Mayr–Patz equation based on the HOMO energy of the nucleophile and the LUMO energy of the electrophile has been presented by Zhuo *et al.* [104]. More recently, Chamorro *et al.* [105] developed an intrinsic (i.e., electronic) relative scale of electrophilicity and nucleophilicity that is the theoretical analog of the Mayr–Patz equation. When the nucleophile N interacts with the electrophile E, the flow of electron density from N to E modulates the electronic interaction between them. The maximum amount of electronic stabilization, ΔE^*_{EN}, occurs at the optimal extent of charge transfer from N to E. Factoring this electron stabilization energy into electronic stabilization energies associated with the individual species E and N gives the relationship shown in Equation 43:

$$\Delta E^*_{EN} = \Delta E^*_{E(N)} + \Delta E^*_{N(E)} \qquad (43)$$

that can be expressed explicitly in the form (Equation 44):

$$\Delta E^*_{EN} = \frac{1}{2}\frac{[\mu_N^2 - 2\mu_N\mu_E]}{(\eta_N + \eta_E)} - \frac{1}{2}\frac{\mu_E^2}{(\eta_N + \eta_E)} \qquad (44)$$

In this equation, μ_N and μ_E are the chemical potentials ($\mu_N > \mu_E$) and η_N and η_E the corresponding hardnesses of N and E, where $\mu \approx -(IP + EA)/2$ and $\eta \approx IP - EA$ can be expressed in terms of the vertical ionization potential ($IP \sim -\varepsilon_{HOMO}$ as the first approximation) and electron affinity ($EA \sim -\varepsilon_{LUMO}$) for each species. For an electrophile immersed in a sea of electrons provided by an ideal donor, its intrinsic or nucleophile-independent electrophilicity ω_E is given by Equation 45:

$$\omega_E = -\frac{1}{2}\frac{\mu_E^2}{\eta_E} \qquad (45)$$

The second term on the right-hand side of Equation 44 is the negative of the relative electrophilic power of E in the presence of the nucleophile N and, upon substitution of

Equation 45, can be written as Equation 46:

$$\omega_{E(N)} = \frac{1}{2}\frac{\mu_E^2}{(\eta_N + \eta_E)} = \omega_E\frac{\eta_E}{(\eta_N + \eta_E)} \tag{46}$$

Employing Equation 46, Equation 44 can then be transformed into a form (Equation 47) that is the electronic analog of the Mayr–Patz equation:

$$\Delta E_{EN}^* = s_{N(E)}[\omega_{N(E)} + \omega_E] \tag{47}$$

where $S_{N(E)}$ is defined by Equation 48:

$$s_{N(E)} = -\frac{\eta_E}{\eta_N + \eta_E} \tag{48}$$

The term $\omega_{N(E)}$, defined in Equation 49:

$$\omega_{N(E)} = \frac{1}{2}\frac{[\mu_N^2 - 2\mu_N\mu_E]}{\eta_E} \tag{49}$$

can be identified as the relative nucleophilicity index of N in the presence of the electrophile E. These theoretically based relationships [105] provide a framework, within the limits of the approximations used to derive them (which include neither solvent nor steric hindrance effects), for analyzing the electronic factors that control the interactions of nucleophiles and electrophiles.

The current status of global and local electrophilicity indices has been exhaustively treated in a review that has been periodically updated [106]. Rezende and Millán [77] have considered the relationship between substituent constants for electrophilicity and Hammett σ constants and between the solvatochromic behavior of dyes and the electrophilicity indices of the donor and acceptor regions of the dyes [107–109].

8 SOLVENT AND SOLUTE PARAMETERS

8.1 Single-Parameter Solvatochromic Probes

Solvation plays a fundamental role in determining reactivity. Polar solvents stabilize and facilitate the development of charge in the transition state, but the energy required to remove solvent from a nucleophile can drastically reduce its nucleophilicity in polar or hydrogen-bonding solvents. Earlier measures of solvent polarity such as Kosower's Z scale, based on the solvent-dependent shifts of the absorption of the charge transfer band of *N*-ethyl-4-methoxycarbonylpyridinium iodide, have been largely supplanted by the much more widespread use of the Dimroth–Reichardt E_T scale of solvent polarity, based on the solvatochromic shifts of the absorption of the pyridinium-*N*-phenolatebetainedye, namely Reichardt's E_T(30) betaine (Scheme 5).

The E_T solvatochromic scale is operationally defined in terms of the absorption maximum (λ_{max}, in nm) of the probe dissolved in the medium of interest, which is converted to the equivalent in kilocalories per mole via the relationship: $E_T(\text{probe}) = 2.859 \times 10^4/\lambda_{max}$.
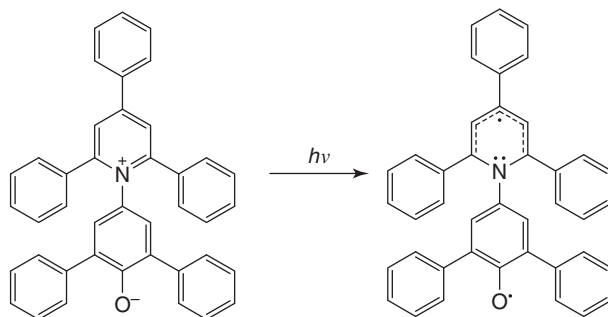
Scheme 5 Reichardt's pyridinium-*N*-phenolate betaine used to develop the E_T scale of solvent polarity on the basis of its absorption maximum [110].

The corresponding normalized dimensionless value, denoted as E_T^N, is defined according to Equation 50:

$$E_T^N = \frac{[E_{T(solvent)} - E_{T(TMS)}]}{[E_{T(water)} - E_{T(TMS)}]} = \frac{[E_{T(solvent)} - 30.7]}{32.4} \tag{50}$$

where $E_{T(Water)}$ and $E_{T(TMS)}$ are the reference E_T values in water and in tetramethylsilane. Extensive compilations of E_T values are available in the literature [110, 111] and E_T and E_T^N values for a selection of commonly used solvents are listed in Table 3.

In protic solvents, both polarity of the solvent and specific hydrogen bonding interactions with the phenolate oxygen of the probe contribute to E_T (see below). In addition to the dependence on the nature and composition of the solvent (solvatochromism), the absorption spectra of pyridinium-*N*-phenolatebetaines are sensitive to temperature (thermochromism), the nature and concentration of added electrolytes (halochromism), and the external pressure (piezochromism) [118], leading to the suggestion that these betaines be called perichromic dyes [110, 119]. The thermochromic and halochromic shifts of the absorption spectrum of the E_T probe have been shown to arise from the complexation of the cation with the phenolate oxygen of the probe [118, 120, 121]. Rezende *et al.* [107–109] have also discussed the relationship between solvent electrophilicity and the solvatochromic shifts of dyes.

Although empirical solvatochromic descriptors such as E_T are useful for developing predictive correlations and estimating effective polarities in nonhomogeneous or unconventional media, the interpretation of such correlations in terms of specific solvent–solute interactions is not always straightforward. For example, E_T values in mixed solvents are sensitive to specific solvation effects, reflecting the fact that absorption spectra are much more sensitive to the local solvation environment around the probe rather than to bulk solvent polarity (see below).

8.2   Lipophilicity Parameters

Lipophilicity provides a measure of the affinity of molecules for a nonaqueous solvation environment as opposed to an aqueous environment. By far, the most widely used lipophilicity parameter for solutes is log $P_{o/w}$, where $P_{o/w}$ is the partitioning coefficient for transfer of the solute from water saturated with *n*-octanol to *n*-octanol saturated with water, defined

TABLE 3 Values of Solvent Parameters for a Selection of Common Solvents

Solvent[a]	Dimroth–Reichardt[b, c]		Kamlet–Taft[b, d]			Catalan[b, e]			Gutmann[b, f, g]		Swain[b, h]		Hansch[i]
	E_T	E_T^N	α	β	π	SA	SB	SSP	AN	DN	A	B	$\log P_{o/w}$
Hexane	31.0	0.01	0.00	0.00	-0.04	0.00	0.06	0.52	0	0	0.01	-0.01	3.90
CCl₄	32.4	0.05	0.00	0.10	0.28	0.00	0.04	0.63	8.6	0	0.09	0.34	2.83
Benzene	34.3	0.11	0.00	0.10	0.59	0.00	0.12	0.67	8.2	0.1	0.15	0.59	2.13
Diethyl ether	34.5	0.12	0.00	0.47	0.27	0.00	0.56	0.69	3.9	19.2	0.12	0.34	0.89
THF	37.4	0.21	0.00	0.55	0.58	0.00	0.59	0.84	8	20	0.17	0.67	0.46
Ethyl acetate	38.1	0.23	0.00	0.45	0.55	0.00	0.54	0.80	9.3	17.1	0.21	0.59	0.73
CHCl₃	39.1	0.26	0.20	0.10	0.58	0.05	0.07	0.79	23.1	4	0.42	0.73	1.97
HMPA	40.9	0.31	0.00	1.05	0.87	0.00	0.81	0.93	10.6	38.8	0.00	1.07	0.28
Acetone	42.2	0.35	0.08	0.43	0.71	0.00	0.48	0.88	12.5	17	0.25	0.81	-0.24
t-Butanol	43.7	0.40	0.28	0.93	0.40	0.15	0.93	0.83	27.1	21.9	0.45	0.50	0.35
DMF	43.8	0.40	0.00	0.69	0.88	0.04	0.29	0.90	16	26.6	0.30	0.93	-1.01
DMSO	45.1	0.44	0.00	0.76	1.00	0.07	0.65	1.00	19.3	29.8	0.34	1.08	-1.35
Acetonitrile	45.6	0.46	0.19	0.40	0.75	0.04	0.29	0.90	18.9	14.1	0.37	0.86	-0.34
Isopropanol	49.2	0.57	0.76	0.84	0.48	0.28	0.83	0.85	33.8	21.1	0.59	0.44	0.05
Methanol	55.4	0.76	0.98	0.66	0.60	0.61	0.55	0.86	41.5	19	0.75	0.50	-0.77
Water	63.1	1.00	1.17	0.47	1.09	1.06	0.03	0.96	54.8	18	1.00	1.00	-1.38
CF₃CH(OH)CF₃	65.3	1.07	1.96	0.00	0.65	1.01	0.01	1.01	88	0	—	—	1.66

The parameters include the Dimroth–Reichart solvatochromic polarity parameter (E_T) and its normalized equivalent (E_T^N), the Kamlet–Taft parameters for solvent hydrogen bond acidity (α/SA), hydrogen bond basicity (β/SB) and solvent dipolarity (π/SSP), the Gutmann solvent acceptor (AN) and donor (DN) numbers, and the Swain solvent Acity (A) and Basity (B). The octanol–water partitioning coefficients ($\log P_{o/w}$) of the individual solvent molecules are also listed for comparison.

[a]THF, tetrahydrofuran; HMPA, hexamethylphosphoramide; DMF, N,N-dimethylformamide; DMSO, dimethylsulfoxide.

[b]Reference [112].
[c]Reference [110].
[d]Reference [113].
[e]Reference [114].
[f]Reference [115].
[g]Reference [116].
[h]Reference [117].
[i]Reference [36].

(Equation 51) as the ratio of the concentrations of the solute in the two phases [36]:

$$P_{o/w} = \frac{[C_{octanol}]}{[C_{water}]} \tag{51}$$

A wide variety of important phenomena correlate with $\log P_{o/w}$, including micellization, solubilization and transport of molecules across biological membranes, and concentration and dispersion of pollutants in the environment [36]. For solutes that ionize, one can use the log of the octanol–water distribution coefficient, $\log D_{o/w}$, where $D_{o/w}$ is defined as the ratio of the concentration of solute in the octanol phase divided by the total concentration of solute (ionized plus unionized) in the aqueous phase buffered at the pH of interest [122].

Fujita *et al*. [123] defined the local lipophilicity parameter π_X for a substituent X as shown in Equation 52 in terms of the difference between the $\log P_{o/w}$ values for compounds with (R–X) and without (R–H instead of R–X) the substituent of interest:

$$\pi_X = \log[P_{o/w}(R-X)] - \log[P_{o/w}(R-H)] \tag{52}$$

Noting that Equation 52 can be rearranged to Equation 53:

$$\log[P_{o/w}(R-X)] = \log[P_{o/w}(R-H)] + \pi_X \tag{53}$$

values of π_X can also be employed to calculate unknown values of $\log[P_{o/w}(R-X)]$ if the value of $\log[P_{o/w}(R-H)]$ is known.

Experimentally, $\log P_{o/w}$ values can be determined directly by measuring the concentration of the solute in the two phases or indirectly by reversed phase thin layer chromatography (TLC) or high-performance liquid chromatography (HPLC) [124]. Alternatively, they can be estimated by summing tabulated hydrophobic fragment constants or group equivalents for all groups present in the molecule [125–127]. The comprehensive compilation of such values by Hansch *et al*. [36] contains several thousand critically evaluated $\log P_{o/w}$ values. The Syracuse Research Corporation (SRC) PHYSPROP database [128] provides online access to the thermophysical properties, including $\log P_{o/w}$ where available, of about 25 000 organic compounds. A number of computer programs have also been developed to estimate $\log P_{o/w}$ from chemical structure. Free estimation programs include the Virtual Computational Chemistry Laboratory and the Estimation Program Interface (EPI) Suite; the former provides online calculation of $\log P_{o/w}$ values with the ALOGPS program [129, 130] and the latter, distributed by the US Environmental Protection Agency (EPA) [131], calculates $\log P_{o/w}$ and a wide range of other properties related to the environmental fate of chemicals. A large, chemically diverse $\log P_{o/w}$ data set has recently been validated for developing and benchmarking such prediction programs [132].

8.3 Hydropathy Scales

Hydrophobic amino acid residues are generally located in the interior of protein domains, whereas hydrophilic amino acid residues are typically found on the exterior surface of proteins. With this in mind, several scales of hydropathy that classify amino acids as hydrophobic or hydrophilic according to the relative hydrophobicity/hydrophobicity of

their side-chain residues have been developed to aid the prediction of protein structure and folding. Although these scales agree qualitatively, there are differences in the relative ordering of the amino acids predicted by the various scales [133]. White and Wimley [134] have developed a whole-residue hydrophobicity scale for amino acids in peptides that has significant advantages over these earlier side-chain hydropathy scales. On the basis of the values of the Gibbs energies for transfer of polypeptides from water to octanol or a phosphatidylcholine (POPC) bilayer interface, the Wimley–White scales include the peptide bond and the side chain. These scales are conveniently available online on the White group website [135] and have been implemented in Membrane Protein Explorer (MPEx), a tool for exploring the topology and other features of membrane proteins by means of hydropathy plots based upon thermodynamic and biological principles [136].

8.4 Kamlet–Taft Solvatochromic Parameters for Solvents

The first multiparameter model for solvent effects was developed by Koppel and Palm, who correlated solvent effects on reactivity using the LSER shown in Equation 54 [137]:

$$\log k_{\text{Rxn}} = \text{const} + gf(\varepsilon) + pf(n_{\text{D}}) + cE + bB \tag{54}$$

where $f(\varepsilon) = (\varepsilon - 1)/(2\varepsilon + 1)$ is a function of the static dielectric constant of the solvent, $f(n_{\text{D}}) = (n_{\text{D}}^2 - 1)/(n_{\text{D}}^2 + 2)$ a function of the refractive index of the solvent, and E and B represent the electrophilic and nucleophilic solvation abilities of the solvent, respectively.

A more general model for treating solvent and solute effects, which made explicit connections between macroscopic solute parameters and microscopic interactions, was developed by Kamlet, Taft, and coworkers [137]. In this model, the overall solute–solvent interaction is factored into the sum of three basic types of contributions: (i) the energy required to create a cavity in the solvent that is the size of the solute, reorganize the solvent around the cavity and insert the solute into the cavity (which results in favorable dispersion interactions between the solute and the surrounding solvent); (ii) the nonspecific interactions between the solute and the surrounding solvent due to either polarizability or solute–solvent dipolarity effects; and (iii) the specific hydrogen bond donor–acceptor interactions, whose strength depends on the solvent/solute hydrogen bond basicity and the solute/solvent hydrogen bond acidity.

On the basis of this model, Kamlet, Taft, and coworkers [137] developed three new solvatochromic descriptors based on solvent-induced shifts of the absorption maxima of probe molecules to represent solvent dipolarity (π^*), solvent hydrogen bond acidity (α), and solvent hydrogen bond basicity (β). In the resultant linear Gibbs energy relationships, these three chemically meaningful descriptors are supplemented by a polarizability parameter δ, which takes the values of 1 (aromatics), 0.5 (halocarbons), or 0 (other molecules), and a cavity term given by the Hildebrand parameter δ_{H}^2 for the cohesive energy density of the solvent. This is the energy required to create a cavity of unit volume in the solvent, which can be calculated from the enthalpy of vaporization of the solvent divided by its molar volume, as shown in Equation 55:

$$\delta_{\text{H}}^2 = \frac{(\Delta H_{\text{vap}} - RT)}{V_{\text{m}}} \tag{55}$$

The effect of solvent on the reaction rate (or equilibrium constant) of interest can then be correlated via the linear Gibbs energy relationship shown in Equation 56:

$$\log k_{Rxn} = const + s(\pi^* + d\delta) + a\alpha + b\beta + h\frac{\delta_H^2}{100} \tag{56}$$

where the coefficients of each of the Kamlet–Taft parameters correspond to the relative influence of each solvent–solute interaction on the reaction rate (or equilibrium position). Addition of the Hildebrand parameter to the Koppel–Palm equation (Equation 54) to include the effects of cavitation makes it formally equivalent to the Kamlet–Taft approach. A clear advantage of the Kamlet–Taft model over other alternative models is that the values of the solvent parameters in Equation 56 have been determined for a much larger number of solvents. Kamlet–Taft α, β, and π^* values for a selection of commonly used solvents are listed in Table 3. In addition to common solvents [112, 113], values of Kamlet–Taft solvent parameters have been reported for a variety of solvents of interest in green chemistry, including ionic liquids as well as several sub- and supercritical solvents [138–143]. Particularly interesting would be the use of Kamlet–Taft parameters [140, 141, 144] to rationalize the unique reactivity of organic molecules in superheated and supercritical water under pressure [144–146].

Three examples illustrate the type of chemical insight that can be obtained from the Kamlet–Taft linear Gibbs energy relationships for solvent effects. Marcus [115] showed that the normalized values of Reichardt's E_T correlated with both π^* and α via Equation 57:

$$E_T^N = 0.01 + 0.36\,\pi^* + 0.47\alpha \tag{57}$$

reflecting the fact, as noted earlier, that Reichardt's solvatochromic dye (Scheme 5) is also a hydrogen bond acceptor. The absence of the cavity term is expected since the size of the solvatochromic dye does not change during absorption of light. The second example is the unimolecular solvolysis of *tert*-butyl chloride, for which Abraham *et al.* [147] reported the relationship shown in Equation 58:

$$\log k_{(t\text{-}BuCl)} = -14.60 + 5.10\pi^* + 4.17\alpha + 0.73\beta + 0.48\frac{\delta_H^2}{100} \tag{58}$$

The relative magnitudes of the coefficients of Equation 58 indicate that the solvolysis is accelerated primarily by solvent dipolarity (π^*) and hydrogen bond donor strength (α), with only minor contributions from solvent cavitation ($\delta_H^2/100$) and hydrogen bond basicity (β). For the Menshutken reaction or S_N2 reaction between triethylamine and ethyl iodide (Scheme 6), Abraham *et al.* [137] found the following relationship (Equation 59) for the Gibbs energy of activation relative to dimethylformamide:

$$\Delta^{\ddagger}G_{solvent} - \Delta^{\ddagger}G_{DMF} = 5.41 - 7.31(\pi^* - 0.18\delta) + 0.54\frac{\delta_H^2}{100} \tag{59}$$

showing that the overriding factor determining the Gibbs energy of activation is the stabilization of the transition state by solvent dipolarity.

As indicated by Equation 57, E_T values are subject to specific solvation effects that can be particularly complicated to model in mixed solvents. El Seoud and coworkers [148–155] have elucidated the individual contributions to the solvatochromism of a
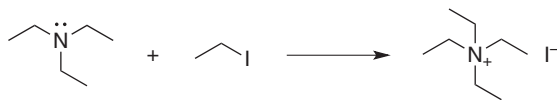
Scheme 6   Menshutken reaction between triethylamine and ethyl iodide.

variety of merocyanine, quinolinium, and pyridiniumbetaine dyes by correlating the $E_{T(Probe)}$ values of these probes with an equation based on Kamlet–Taft parameters and the octanol–water partitioning constant to take into account the sensitivity of the merocyanine probes to solvent lipophilicity, as shown in Equation 60:

$$E_{T(probe)} = \text{const} + a\alpha + b\beta + s(\pi^* + d\delta) + p\log P_{o/w} \tag{60}$$

Studies of the thermo-solvatochromism of these probes in binary aqueous-organic solvent mixtures led to the development of a general model based on preferential solvation of the probe by the organic component and, even more efficiently, by the organic component hydrogen bonded to water. These concepts have a direct bearing on the development of suitable QSRR treatments of reactivity in mixed solvents, which can be particularly complex when these solvent mixtures exhibit microheterogeneity.

8.5   Catalán's SSP, SB, and SA Solvent Parameters

Catalán [156] has developed an alternative set of solvatochromic probe-based descriptors: the solvent polarity/polarizability (SPP) scale, the solvent hydrogen bond basicity (SB) scale, and the solvent hydrogen bond acidity (SA) scale. Values of these three parameters are available for more than 200 solvents [111, 112, 114], some of which are listed in Table 3. Catalán has related his parameters to other descriptors and has applied them to selected reactions and equilibria [156]. For comparison with the Kamlet–Taft approach (Equation 57), Catalán's equivalent linear Gibbs energy relationship for E_T^N is shown in Equation 61:

$$E_T^N = -0.31 + 0.62\text{SPP} + 0.77\text{SA} + 0.12\text{SB} \tag{61}$$

The major difference between this equation and that of the Kamlet–Taft approach is the presence of the SB term. Catalan's linear Gibbs energy relationships for the solvolysis of *tert*-butylchloride are particularly interesting. In 27 pure solvents, he obtained the relationship in Equation 62:

$$\log k_{(t\text{-BuCl})} = -19.85 + 10.02\text{SPP} + 8.03\text{SA} + 1.84\text{SB} \tag{62}$$

When solvolysis data for an additional 120 binary aqueous-organic solvent mixtures were included, the almost identical relationship shown in Equation 63 was obtained:

$$\log k_{(t\text{-BuCl})} = -20.07 + 10.62\text{SPP} + 7.89\text{SA} + 1.71\text{SB} \tag{63}$$

These results suggest that specific solvation effects are not particularly important for this reaction. For the Menshutken reaction between triethylamine and ethyl iodide, the Catalán approach gave the relationship in Equation 64:

$$\log k_{(TEA/EtI)} = -4.07 + 8.84\text{SPP} + 1.90\text{SA} \tag{64}$$

As in the Kamlet–Taft approach (Equation 59, the importance of solvent (di)polarity effects on the reaction rate is manifested in Equation 64 via the coefficient of the SSP term.

8.6   The Grunwald–Winstein Equation

The predecessor to the multiparametric solvent parameter equations discussed earlier was the Grunwald–Winstein linear Gibbs energy relation [157, 158], shown in Equation 65:

$$\log\left(\frac{k_S}{k_0}\right) = mY + \text{const} \tag{65}$$

This one-parameter equation was used to correlate rate constants for the solvolysis of the substrate in the solvent of interest and in the reference solvent (ethanol/water 80/20), both measured at 25 °C. The slope m was assigned a value of unity (with the constant term = 0) for the unimolecular solvolysis of *tert*-butyl chloride, which created a set of descriptors Y for the relative ionizing power of the solvent. As shown in Equations 58 and 62, the rate of solvolysis of *tert*-butyl chloride is accelerated primarily by the dipolarity and hydrogen bond donor strength of the solvent, suggesting that these are the dominant factors codified by the Y values. Streidl and Mayr [159] have determined Y values for aprotic solvents. The original Grunwald–Winstein Y values are now known to be adequate only when chloride is the leaving group. For other leaving groups, the rate constants for solvolysis of 1- or 2-adamantyl substrates, where X is the leaving group of interest, have been used to determine leaving group-dependent Y_X values for the ionizing power of solvents.

The basic Grunwald–Winstein equation has also been expanded to include solvent nucleophilicity, N_T, and an aromatic ring parameter, I [158], as shown in Equation 66:

$$\log\left(\frac{k_S}{k_0}\right) = mY + \ell N_T + hI + \text{const} \tag{66}$$

where ℓ and h are the corresponding sensitivity parameters. The preferred scale for the N_T values is that based on the nucleophilic attack of the solvent of interest at the methyl group of the *S*-methyldibenzothiophenium ion [160].

8.7   Gutmann's Solvent AN and DN

The Gutmann donor (DN) and acceptor (AN) numbers are classical examples of parameters designed to measure the strength of the Lewis basicity or Lewis acidity of solvents, respectively [115]. The donor number was originally based on the heat of reaction between $SbCl_5$ and an equimolar amount of the solvent molecule in $ClCH_2CH_2Cl$. Subsequently, ^{23}Na NMR chemical shifts of $NaClO_4$ dissolved in the solvent were found to be a more convenient method of determining the DN. Marcus [115, 161] normalized the donor numbers by dividing them by 38.8, the DN for hexamethyphosphoramide (HMPA); the normalized donor numbers (DN^N) were found to correlate reasonably well with Kamlet–Taft's β values. The values of AN are based on the ^{31}P-NMR chemical shift of triethylphosphine oxide dissolved in the solvent of interest. Marcus [115] also showed that the Gutmann AN parameter correlated well with a linear combination of Kamlet–Taft's α and π^* values. The values of the DN and AN are tabulated in the literature [111, 115, 116] and online [112]. Recently, Schmeissser *et al*. [162] reported Gutmann donor and acceptor numbers for ionic liquids

and found correlations between their DNs and Kamlet–Taft β values, but no correlation between their ANs and Kamlet–Taft α values.

Solvatochromic probes have also been used to determine the donor numbers of anions [163] and the acceptor numbers of cations [164, 165] in solution. Measurements in different solvents showed that the apparent anion donor numbers decreased with increasing solvent AN and that the apparent cation acceptor numbers decreased with increasing solvent DN.

8.8   Solvent Acity and Basity

Swain *et al.* [117, 166] developed a two-parameter scale (Equation 67) for the anion and cation solvation tendencies of solvents, referred to as "Acity" and "Basity," respectively.

$$\log k = a \ \text{Acity} + b \ \text{Basity} + \text{const} \tag{67}$$

The Acity and Basity descriptors for 55 solvents and 6 aqueous solvent mixtures were determined by application of this two-variable equation to a large number of reactions, and the values were normalized by assuming that Acity = Basity = 0 for *n*-heptane and Acity = Basity = 1 for water, with the additional conditions that Acity = 0 for HMPA and Basity = 0 for trifluoroacetic acid. Values of the Acity and Basity of a variety of solvents are tabulated in the literature [111, 115] and online [112]. Several representative values are given in Table 3. Marcus [115] showed that there are reasonable correlations between Basity and Kamlet–Taft's π^* values, with a small contribution from Kamlet–Taft's β (Basity = $0.04 + 0.94\pi^* + 0.035\beta$). The Acity correlates well with Gutmann's AN and with a linear combination of Kamlet–Taft's α and π^* descriptors (Acity = $0.03 + 0.64\alpha + 0.25\pi^*$).

8.9   Abraham Solute Parameters

Abraham proposed the use of LSERs based on the solute parameters V, L, E, S, A, and B in order to gain detailed chemical insight into the relative contributions of the solvent–solute interactions involved in partitioning phenomena and reactivity [167–169]. As shown in the Kamlet–Taft solvent parameter approach (Figure 9), the Gibbs energy for transferring a solute from one phase to another is factored into three basic components: (i) the energy required to create a solvent cavity of the size of the solute; (ii) the contributions from non-specific (polarizability and dipolar) interactions upon filling the cavity with the solute; and (iii) specific hydrogen bond donor–acceptor interactions between solute and solvent. The solute-specific descriptors *V* and *L* are related to the energy required to form the cavity in the solvent and to the van der Waals interactions between an alkane solute and the solvent cavity. The molar volume of the solute or *V* (in $dm^3 \ mol^{-1}/100$) can be calculated directly from the structure of the solute [170], while *L* is the logarithm of the experimental partitioning coefficient of the solute at 298 K between the vapor phase and hexadecane. The solute descriptors for the nonspecific interactions are *E*, the solute molar refraction in excess of that of an alkane of equivalent molar volume, and *S*, the solute dipolarity. The descriptors A and B are, respectively, the overall hydrogen bond acidity and basicity of the solute. For aniline derivatives and substituted pyridines, the value of *B* can be different for water-organic solvent partitioning phenomena and should be replaced by the descriptor B^0.

An early Abraham review, in which these descriptors were still denoted as V_X, log L^{16}, R_2, π_2, $\Sigma\alpha_2$, $\Sigma\beta_2$, and $\Sigma\beta_2{}^0$, respectively, reported solute parameters for some 500+ solutes
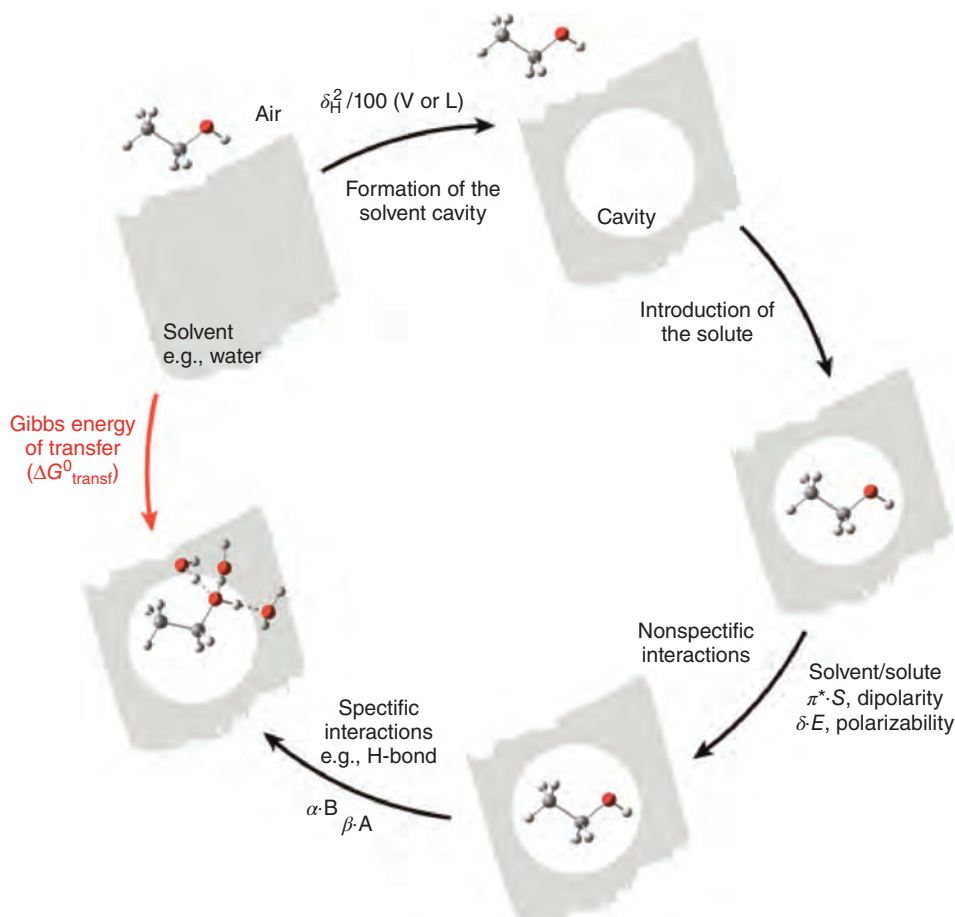
$$\Delta G^0_{transf} = const + a_1, \pi^* \cdot S + a_2 \delta \cdot E + a_3 \alpha \cdot B + a_4 \beta \cdot A + a_5 (\delta_H^2/100)(V \text{ or } L)$$

Figure 9   Thermodynamic cycle for the Gibbs energy of transfer (ΔG^0_{trans}) from the vapor phase to solution factored into individual terms for solvent cavitation and the general dispersion interactions upon insertion of the solute into the cavity; the nonspecific solvent–solute dipolar interactions and polarizability; and the specific solvent–solute hydrogen bond interactions (see the text for the definition of the solvent and solute parameters). The general solvent–solute linear solvation free energy relationship (LFER) indicated in the figure reduces to the Kamlet–Taft LSER, Equation 56, for a single solute partitioning into multiple solvents and to the Abraham LSERs, Equations 68 and 69, for partitioning between two phases and for partitioning of multiple solutes into a single solvent, respectively.

[167]. Experimental measurement of E, A, B, and B^0 by chromatographic methods has been reviewed by Poole *et al*. [169], and quantum chemical descriptors calculated from molecular structure have been used to estimate the Abraham parameters [171]. Currently, Abraham solute descriptors are available for approximately 3700 solutes and these have been conveniently compiled by the Helmholtz Centre for Environmental Research (UFZ) in their online UFZ-LSER database [172].

A limitation of the V values employed by Abraham is that they are identical for constitutional isomers. Values of V corrected for branching have been developed for the constitutional isomers of acyclic alkanes [173] and shown to improve LSERs involving these compounds [173, 174]. Subsequently, van Noort *et al*. [175] showed that V values corrected for alkyl branching also improve the LSERs for non-hydrocarbon molecules. An important consequence of using branching-corrected values of V is that some of Abraham's A and B descriptors obtained from experiment may also have to be corrected, but this has been done for only a limited number of compounds [175].

Solute properties (SPs) involving condensed phases are usually correlated with an LSER of the type shown in Equation 68:

$$\log SP = \text{const} + eE + sS + aA + bB + vV \tag{68}$$

whereas those involving vapor-phase-condensed phase partitioning are correlated with an LSER of the type shown in Equation 69:

$$\log SP = \text{const} + eE + sS + aA + bB + \ell L \tag{69}$$

where log SP is a Gibbs-energy-related solute property such as the log of a partitioning coefficient. The descriptors are approximately normalized, which means that the magnitudes and signs of the multiple regression coefficients e, s, a, b, v, and ℓ provide a measure of the relative importance and direction of influence of each of these individual solute–solvent interactions. For processes such as transfer of the solute between two condensed phases, the coefficients of Equation 68 represent the difference between the contributions of the two phases. The physical–chemical significance of the constant term has been analyzed in the light of solvation thermodynamics by van Noort [176].

These equations have been applied to a wide range of phenomena, including solute partitioning in polar and nonpolar solvents, biological systems, micelles, ionic liquids, binary solvent systems, and supercritical fluids, as well as gas–liquid partitioning equilibria; they have also been employed to explain the toxicity of organic molecules to aquatic organisms, enthalpies of solvation, and so on [168, 177, 178]. In all of these applications, the underlying assumption is that the average solvation environment is not solute dependent. Thus, the Abraham approach should be less satisfactory when selective solvation becomes important, such as in mixed aqueous-organic solvents, or when solutes occupy multiple solubilization environments.

A particularly interesting LSER is Equation 70 for the partitioning coefficient of solutes between octanol and water:

$$\log P_{o/w} = 0.08 + 0.58E - 1.09S + 0.03A - 3.40B + 3.81V \tag{70}$$

This equation shows that an increase in solute size favors partitioning to the octanol phase, which has a much lower cohesion energy than water, whereas greater dipolarity and hydrogen bond basicity of the solute favor partitioning to the aqueous phase. This LSER thus provides a chemically meaningful perspective on the specific molecular features that favor or disfavor octanol–water partitioning. Analogous effects of solute size and hydrogen bond basicity are observed for the solubilization of organic molecules in sodium dodecyl sulfate (SDS) micelles in aqueous solution [179, 180]:

$$\log K_{mic/w} = 0.08 + 0.58E - 1.09S + 0.03A - 3.40B + 3.81V \tag{71}$$

Thus, despite decades of speculation that solutes of different hydrophobicities solubilize in different regions of micelles (hydrocarbon core and micelle–water interface), an LSER that implicitly assumes the existence of a single micellar solubilization environment works remarkably well for all types of neutral organic solutes.

The LSER shown in Equation 72 for the gas-phase-water partitioning of organic molecules [181]:

$$\log L_w = -0.994 + 0.577E + 2.549S + 3.813A + 4.841B - 0.869V \qquad (72)$$

indicates that solute dipolarity and hydrogen bond acidity and basicity all favor solubilization in the aqueous phase, but that the partitioning is only modestly sensitive to solute molar volume. The lack of a strong influence of solute size was attributed to a compensation between two large Gibbs energy contributions, that is, the unfavorable energy of cavity formation in water and the favorable van der Waals interactions between the solute and the solvent cavity, both of which increase with solute molar volume.

Although the Abraham descriptors are formally solute parameters, they have been shown to correlate several properties of condensed phases, such as the surface tension of organic liquids [182] and the interfacial adhesion between water and organic liquids, $W_{o/w}$ [183], with chemically reasonable coefficients. The LSER for the surface tension of pure organic liquids is dominated by the nonspecific interaction terms in E and S with only a minor contribution from hydrogen bonding (included as the square root of the product AB to include only solvents capable of self-hydrogen bonding) [182]. This is consistent with the tendency of molecules at the surface of the liquid to maintain their hydrogen bonds when possible. On the other hand, the LSER for the Gibbs energy required to separate a bulk organic phase from contact with water, $W_{o/w}$, was found to be strikingly similar to that for gas-phase-water partitioning of organic solutes (Equation 72). This implies that the interactions between an organic solute and the surrounding aqueous cavity parallel, at the molecular level, the interactions present at the bulk organic liquid–water interface [183]. This is in line with both molecular dynamics simulations and interfacial vibrational spectroscopic results that show that there are free or "dangling" OH bonds at the air–water interface and around the nonpolar portions of organic solutes dissolved in water, but not at water–hydrophilic interfaces or around the more polar groups of solutes [184–188].

9   TEMPERATURE-INDEPENDENT LINEAR GIBBS ENERGY RELATIONSHIPS

The dependence of the Hammett ρ value on the inverse of the temperature was mentioned previously (Equation 22) in the context of the isokinetic temperature. For QSRRs with more than one descriptor, it is reasonable to assume that the sensitivity of the reaction or equilibrium to each descriptor X_i might have its own temperature dependence $\rho_i(T)$ and its own effective compensation temperature β_I at which $\rho_i(\beta_I) = 0$. With this in mind, one can generalize Equation 22 for the case of a quantitative structure–reactivity relationship (QSRR) containing **n** descriptors with distinct temperature dependences as follows:

$$\log k_{Rxn} = \sum_{i=0}^{n} \rho_i(T)X_i = \sum_{i=0}^{n} c_i \left(1 - \frac{\beta_i}{T} \right) X_i = \sum_{i=0}^{n} \left(c_i - \frac{c_i{}'}{T} \right) X_i \qquad (73)$$

where c_i and $c_i{}'$ $(=-c_i\beta_i)$ are regression coefficients and $i=0$ represents the constant term. Equations of this type have indeed been shown to account for the temperature dependence of several partitioning equilibria [168, 189–191].

10   QSPR AND QSRR FOR SUPRAMOLECULAR INCLUSION COMPLEXES

Given the enormous variety of supramolecular inclusion complexes that have been investigated, meaningful quantitative structure–property relationships (QSPR) and, in particular, QSRRs for them are relatively scarce. Thus, the discussion of this topic focuses mainly on the inclusion complexes of cyclodextrins and the cucurbiturils (Scheme 7), with only brief mention of other systems. In contrast to the cucurbiturils, the binding of organic molecules to cyclodextrins exhibits quite good enthalpy–entropy compensation [192]. Despite numerous attempts to generate QSPRs for binding of solutes to cyclodextrins, most of the correlations with descriptors such as Abraham parameters or quantum chemical or structural variables are based on a relatively limited number of solutes (<250 and usually much fewer) that are often of similar structural types [193–198]. QSPR modeling of the stability constants of crown ether complexes is also similarly limited in scope [193]. The log of the stability constants for the insertion of disubstituted dibenzylammonium cations into the crown ether dibenzo[24]-crown-8 were shown to correlate linearly with Hammett's σ values [199]. The Gibbs energy of binding of pyrene to a macrocyclic cyclophane was found to correlate linearly with Reichardt's E_T values in 17 organic solvents and solvent mixtures and water [200, 201].

Among the few reported QSRRs for the reactivity of molecules incorporated in the cavities of cyclodextrins and cucurbiturils, D'Souza and Bender [202] obtained the "world's worst Hammett plot" for the hydrolysis of m- and p-substituted phenyl acetates. This was attributed to substituent-dependent differences in the orientation of binding of the substrate in the inclusion complex (substituent inside or outside the cavity) that outweighed any differences in electronic effects. Rate constants for hydrolysis of p-substituted phenyl phenylacetates were correlated with Hammett's sigmas and Hansch's π values [203], but the number of substrates (five) is insufficient for a two-parameter fit (see below). The S_N1 solvolysis of 1-bromoadamantane incorporated into β-cyclodextrin points to a cavity with an effective solvent ionizing power similar to that of 50% ethanol : water, whereas similar measurements with cucurbit[7]uril indicated an ionizing power of the cucurbituril cavity more like that of 60% ethanol : 40% water [204]. A Hammett plot of the solvolysis of monosubstituted benzoyl chlorides in water using σ^+ values shows a break at about $\sigma^+=0.5$; this break reflects a change from a dissociative mechanism for electron-donating and weak-to-moderate electron-withdrawing substituents ($\rho=-2.6$) to an associative mechanism for strongly electron-withdrawing substituents ($\rho=+1.4$). In the cavity of dimethyl-β-cyclodextrin, the solvolysis is inhibited by about two orders of magnitude, but the Hammett plot still shows a change from a dissociative ($\rho=-2.6$) to an associative mechanism ($\rho=+3.3$) at around $\sigma^+=0$. In contrast, in the cavity of cucurbit[7]uril, a single linear Hammett correlation is observed as a function of σ^+, with a slope ($\rho=-3.1$) similar to that in a good ionizing solvent (e.g., 97% trifluoroethanol), indicating that the associative mechanism is inhibited in the cucurbituril cavity [204].
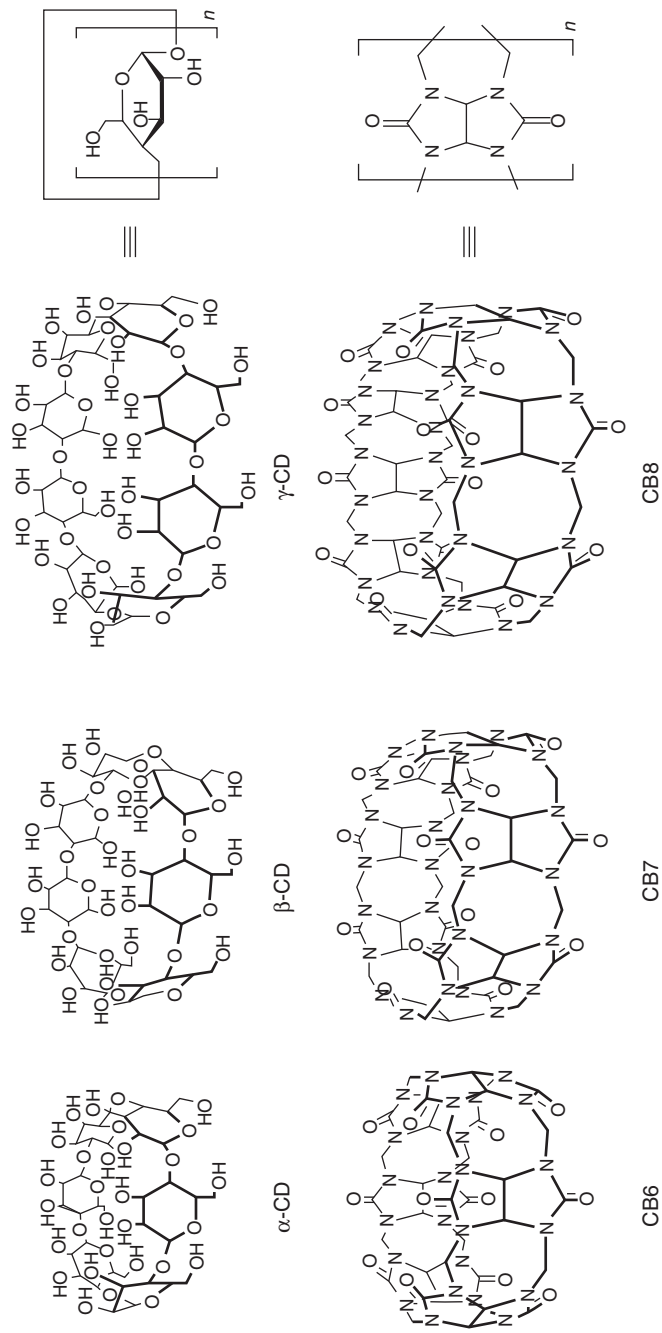
Scheme 7 Structures of the cyclic oligomers α-, β-, and γ-cyclodextrin and the analogous cucurbit[n]urils with $n = 6$, 7, and 8 monomer units, indicating the internal cavity available for the inclusion of solutes.

11   APPLICATIONS OF QSAR

The origins of an understanding of the relationship between the molecular structure of organic molecules and their biological activity date back to the end of the nineteenth century [205]. The modern field of QSARs, however, dates to 1964, with the advent of two approaches, the Free–Wilson method and the Hansch model. In both methods, molecular structure-related descriptors and/or physical–chemical properties are correlated with $\log (1/C_{BioAct})$, where C_{BioAct} is some function of the biological or biochemical activity of interest, such as LD_{50} (the 50% lethal dose) or IC_{50} (the concentration that results in 50% inhibition of an enzyme activity).

The Free–Wilson method [206] uses descriptors that simply indicate whether a given substituent is present or absent at certain positions in the molecule. If the substituent is present, the corresponding indicator variable $\{S_i\}$ for that substituent is assigned a value of unity and if absent a value of zero is employed. The resulting linear QSAR is of the form shown in Equation 74:

$$\log \left(\frac{1}{C_{BioAct}} \right) = \sum a_i \{S_i\} + \mu \tag{74}$$

where μ is the activity of a reference compound such as the unsubstituted analog. The coefficients a_i then provide a measure of the effect and relative importance of a given substituent at each individual position of the molecule on the biological activity. A particularly good example illustrating the result of a Free–Wilson analysis was presented by Kubinyi [207] for the antiadrenergic activity of mono- and disubstituted N,N-dimethyl-α-bromophenethylamines (Scheme 8) with substituents X and/or Y equal to H, F, Cl, Br, I, or Me at the m- and p-positions. Free–Wilson analysis of the antiadrenergic activity gave the QSAR shown in Equation 75:

$$\log \left(\frac{1}{C_{BioAct}} \right) = -0.30\{m\text{-}F\} + 0.21\{m\text{-}Cl\} + 0.43\{m\text{-}Br\} + 0.58\{m\text{-}I\}$$
$$+ 0.45\{m\text{-}Me\} + 0.34\{p\text{-}F\} + 0.77\{p\text{-}Cl\} + 1.02\{p\text{-}Br\}$$
$$+ 1.43\{p\text{-}I\} + 1.26\{p\text{-}Me\} + 7.82 \tag{75}$$

where the constant term refers to the unsubstituted compound (X=Y=H) and the variables (S_i) for each substituent are replaced by the bracketed terms indicating the {substituent and position}.
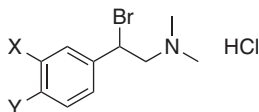
Scheme 8 N,N-Dimethyl-α-bromophenethylamine hydrochloride.

On the positive side, the Free–Wilson approach is model independent, and the indicator variables are easily related to molecular structure. The limitations of this method include the requirements that a minimum of two different positions of the molecule must be chemically modified and that structural features should appear more than once in the data

set. In addition, the number of indicator variables can often be relatively large compared to the number of compounds in the data set, which can complicate the statistical analysis (see below). For example, Equation 75, which contains 10 different S_i values, was based on data for 22 compounds out of a possible 56 different combinations. Moreover, a Free–Wilson QSAR cannot predict the properties of compounds with substituents that were absent from the data set [207]. A Free–Wilson-like local QSAR model that does potentially permit predictions for substituents not present in the original data set has however been proposed by Chen *et al.* [208].

In contrast, the Hansch approach [209] is based on a simplified model, which assumes that biological activity potentially requires passage of the molecules of interest through membrane barriers, followed by their interaction with some final target or active site at which the activity is produced. In its most general form, a QSAR derived from a Hanch analysis can be expressed in the form of Equation 76 [93]:

$$\log\left(\frac{1}{C_{\text{BioAct}}}\right) = -a(\log P_{\text{o/w}})^2 + b(\log P_{\text{o/w}}) + \sum c_i\sigma_i + \sum d_i\delta E_S + \text{const} \qquad (76)$$

The quadratic term with a negative sign was proposed to take into account the parabolic dependence on $\log P_{\text{o/w}}$ of transport through membranes [210]. The rationale for such a dependence is that compounds that are too water soluble will pass through membranes with difficulty, while compounds that are too lipophilic will be retained by membranes. The remaining terms then serve to model the hydrophobic, electronic, and steric components of the interaction of the compounds with the final biological target. Using Hansch's π_X hydrophobic constants for the substituents rather than $\log P_{\text{o/w}}$ values for the compounds as a whole, the Hansch-type QSAR [207] for the same compounds employed in the Free–Wilson analysis (Equation 75) is given by Equation 77:

$$\log\left(\frac{1}{C_{\text{BioAct}}}\right) = 1.15\ \pi_X - 1.46(\sigma_p^+ + \sigma_m^+) + 7.82 \qquad (77)$$

Thus, the QSAR derived from a Hansch approach is not only much simpler with substantially fewer variables but also has the potential of predicting the activity of substituents not present in the original data set from their values of π_X and σ^+ (as long as these fall within the range of descriptor values used to develop the equation). Kubinyi [211] introduced the use of a mixed approach that combined the best features of the Hansch and Free–Wilson models into a single QSAR.

Although these seminal approaches are applicable to the optimization of the biological activity for compounds of a given basic structural type, a more general problem is how to develop QSARs for biological activities such as toxicity or biodegradability for compounds of quite different chemical structures. One promising possibility for the development of chemically meaningful QSARs is to resort to compound-specific parameters such as Abraham solute parameters. Indeed, Abraham solute parameters have been used to correlate data for permeation of solutes through membranes and through the blood–brain barrier, for the anesthetic activity of organic molecules and for toxicity to a variety of aquatic organisms [44, 167, 211–216]. An interesting application is the prediction of the threshold for the detection of the odor of volatile organic compounds by humans. Thus, the log of the odor detection threshold (ODT) for nearly 200 compounds correlated reasonably well with the LSER given in Equation 78 [215]:

$$\log\left(\frac{1}{\text{ODT}}\right) = -1.826 + 0.88\text{E} + 0.408\text{S} + 0.999\text{A} + 2.196\text{B} + 0.578\text{L} + 4.065\text{M}$$
$$+ 1.805\text{AL} + 1.424\text{AC} + 1.290\text{EU} \tag{78}$$

which shows that the five Abraham solute parameters (E, S, A, B, and L) plus the four indicator variables (which have the value of one if the structural feature is present and are zero otherwise) for mercaptans (M), aldehydes (AL), acids (AC), and unsaturated esters (UE) all lower the ODT value.

In many situations of practical interest, however, the knowledge of the system may be too limited to permit a rational choice of variables and/or reliable values of the classical electronic, steric, and solvent/solute descriptors discussed earlier may not be readily available. This has led to a growing tendency to develop QSARs and QSPRs based on constitutional, topological, geometrical, electrostatic, and quantum chemical parameters that can be calculated directly from molecular structure [217–222]. Since computer programs can calculate several thousand such parameters, the number of potential structure-based descriptors can approach or exceed the number of data points. Molecular descriptor [223] is a free online resource that provides tutorials, benchmark data sets, information on books, and journals related to molecular descriptors and links to other online tools and resources. The Molecular Descriptors Data Base (MOLE db) is a free online database composed of more than 1100 molecular descriptors calculated for over 230 000 molecules [217, 224]. A widely used program is Dragon, which can calculate nearly 4900 molecular descriptors of 29 different types from molecular structure, and is available in both a commercial version [225] and an online version called E-Dragon [129].

Most of the molecular descriptors used to date are defined in the recent book by Todeschini *et al.* [220]. However, extracting a valid "optimal" model from this plethora of putative descriptors and QSRR models requires more rigorous mathematical techniques [226–228]. An informational-theoretical analysis based on the Akaike information criterion (AIC) appears to be a particularly robust approach for avoiding model selection bias [229, 230]. In addition, there has been growing interest in the use of 3D (and higher) QSAR approaches [205, 231], which are beyond the scope of this chapter.

An important stimulus to QSAR research has been provided by the European Union regulations for the Registration, Evaluation, Authorisation, and Restriction of Chemicals (REACH), which mandate *the use of alternative test methods, suitable for the assessment of health and environmental hazards of chemicals, wherever possible* in order to avoid animal testing [232]. In this context, appropriately validated QSPR models for predicting chemical properties and QSAR models for assessing chemical risk and exposure and the consequences for health and safety are viewed as acceptable options. Chapter 7R of the European Chemicals Agency Report "Guidance on information requirements and chemical safety assessment" details specific aspects of currently accepted modeling objectives within the scope of REACH [232]. The European Union Institute for Health and Consumer Protection (IHCP) provides an online database of information on QSAR models and a series of downloadable computational tools for analyzing the environmental and toxicological effects of chemicals [233]. The Organisation for Economic Co-operation and Development (OECD)'s QSAR project [234] has developed the QSAR Toolbox, a software application for assessing the hazards of chemicals, as well as a series of guidance documents and five basic principles for the validation of QSAR models.

The North American Free Trade Agreement (NAFTA) has similar QSAR-based regulatory initiatives that are being developed in collaboration between US and Canadian

governmental agencies. The Technical Working Group on Pesticides has recently released detailed guidelines for the development of QSARs [235]. The US EPA also makes available information and software, such as a guide to molecular descriptors for the software TEST [236] for estimating the toxicity and physical properties of organic chemicals based on their molecular structure.

The PaDEL-DDPredictor is an open-source software for calculating the pharmacodynamics, pharmacokinetics, and toxicological properties of chemical compounds [237]. QSARINS (QSAR-INSUBRIA) is a software for the development and validation of multiple linear regression QSAR models using least squares and a genetic algorithm for variable selection [227].

12 QUANTITATIVE ION CHARACTER–ACTIVITY RELATIONSHIPS (QICAR)

A variant of QSAR known as quantitative ion character–activity relationships (QICARs) permits one torationalize and predict the bioactivity of metal cations [238]. Along with QSAR data [239–242], QICAR studies have been shown to be useful in the context of coordination, bioinorganic, and organometallic chemistry. The important factors in QICAR are the physical–chemical properties of the metalcation, the nature of the biological system, and the metal cation–ligand interaction [243]. Particularly useful parameters for the ligand-cation binding tendency are the covalent index and the softness index [238]. The covalent index quantifies the relative importance of covalent over ionic interactions as codified by two fundamental characteristics of ions: their electronegativity and Pauling ionic radius. The softness index separates metal ions into three groups: *hard ions*, which prefer to bind to either oxygen or nitrogen, such as Li(I), Na(I), K(I), Cs(I), Ca(II), Mg(II), Sr(II), Ba(II), Fe(III), and Al(III); *soft ions*, which prefer to bind to sulfur, such as Hg(I), Cu(I), Ag(I), Cd(II), and Pt(II); and *borderline ions*, which form complexes with oxygen, nitrogen, and sulfur, which include Co(II), Fe(II), Ni(II), Cu(II), Zn(II), and Pb(II). Parameters for the "polarizing power" of ions have also been employed to determine metal ion binding to biological ligands [238, 244]. QICARs have been used to predict the toxicity of metal cations to organisms ranging from bioluminescent fungi [245] to seaweeds [246].

13 CONSTRUCTING AND VALIDATING QUANTITATIVE RELATIONSHIPS

The construction and validation of quantitative molecular structure-based relationships involves a number of practical aspects. In the ideal case: (i) each descriptor belonging to a given set of descriptors should be related in a clear-cut manner to one or more of the individual factors that contribute to the Gibbs energy change; (ii) it should be possible to assign to each descriptor a relative magnitude and sign, preferably scaled or normalized in an appropriate manner so that it is comparable to the other descriptors of the set; (iii) when possible, absolute scales that are properties of the individual molecule or of a fixed reference state are better than relative scales that depend on a variable reference state; (iv) each set of descriptors of a given type should be statistically independent of other sets of descriptors used in the correlation; (v) the set or a subset of the descriptors should correlate with one or more physical properties related to reactivity in a statistically robust manner, with appropriate verification of the presumed model (this requires proper use of statistics);

(vi) the best-fit regression coefficients for each descriptor in the resultant linear Gibbs energy relationship should then correspond to the desired sensitivity factors, the magnitude and sign of which indicate the contribution and the relative contribution of each type of descriptor to the overall property of interest; and (vii) the combination of descriptors and sensitivity factors should be chemically meaningful and contribute positively to the development of mechanisms consistent with the observed reactivity patterns.

Dearden *et al*. [247] and Le *et al*. [248] have discussed the potential pitfalls that should be avoided in constructing QSARs and QSPRs. Before attempting correlations, attention must be paid to the quality of the dataset, certifying that it does not contain duplicate, incorrect, or irrelevant data. The descriptors must be checked for incorrect values and for appropriate scaling so that the coefficients in the final linear Gibbs energy relationship express the relative influence of each descriptor. In addition, descriptors should preferably be relatable in a straightforward way to some feature of molecular structure or molecular property, making the final equation amenable to a molecular or mechanistic interpretation.

As the number of variables or different types of descriptors increases in the equation, the number of data points that is required also increases. As a rule of thumb, one should have a minimum of at least five data points for every descriptor in the proposed linear Gibbs energy relationship to avoid overfitting of the data [137, 158]. Thus, for a five-descriptor equation like the Kamlet–Taft equation, data for the process or reaction of interest would have to be obtained in a minimum of 25 solvents. At the same time, however, the experimental data must explore an adequate range of values for each of the descriptors. For example, in choosing the 25 solvents for a linear Gibbs energy relationship based on the Kamlet–Taft equation, one must chose solvents with a wide range of hydrogen-bond basicities and acidities, dipolarities, polarizabilities, and Hildebrand parameters. Once the solvents are chosen, it is then necessary to ensure that there is no appreciable collinearity between the descriptors, that is, that the descriptors are not linearly correlated with each other. In the example of the Kamlet–Taft equation, if any two of the five descriptors π^*, δ_H^2, β, α, or ρ are linearly dependent (e.g., correlation coefficients >0.5 for pairwise linear regressions) for the 25 chosen solvents, one of these two descriptors is no longer a truly independent variable and should be excluded.

Assuming that these criteria are met for the experimental data set and the descriptors, the results of the multiple linear regression must be subjected to standard tests to verify the statistical significance of the best-fit regression coefficients that multiply each descriptor in the final equation. This is particularly important given the current facility of performing multiple linear regression analysis to fit data to multiparametric equations. Most of the standard statistical tests, such as the correlation coefficient (R), the coefficient of determination (R^2), the sum of squares of the residuals, the mean square error, or the F values, provided by such programs are merely indicative of the goodness of the data fitting to the mathematical equation [249]. The goodness-of-fit or statistical significance of the regression coefficients can be judged based on the individual standard errors and the partial F values of the coefficients. At this stage, outliers or data points that are clearly well outside the expected error limits become apparent and demand special attention. Omission of such points from the model should be made explicit in any publication and preferably justified.

Goodness-of-fit to a data set does not, however, necessarily mean that the assumed model is either robust or a valid linear Gibbs energy relationship. This requires an additional validation step in which the assumed linear Gibbs energy relationship is used to predict the observed results for data that were not employed to establish the relationship itself. The simplest procedure is leave-one-out cross validation in which data points are

iteratively excluded one-by-one from the data set and the value of the excluded point predicted based on the remaining members of the data set. If the data set is large enough, it can be pseudo-randomly divided into two subsets, a training set and a test subset. The training subset is used to develop the mathematical model and the predictive power of the model is then tested against the test subset of the data. In either case, comparison of the sum of the errors of the predictions (PRESS) to the sum of errors of the data used to develop the model provides the cross-validated coefficient of determination (symbols R^2_{CV} or Q^2). The domain of $Q^2 \leq 1$ and the optimal model is that which includes only the most relevant descriptors with the best predictive value. On the other hand, values of Q^2 less than zero indicate that the proposed model is actually worse than no model at all. Brief but lucid discussions of these goodness-of-fit and validation parameters can be found online or in the recent literature [249–254].

ACKNOWLEDGMENTS

REFERENCES

1. Carroll, F.A. (2010) *Perspectives on Structure and Mechanism in Organic Chemistry*, 2nd edn, John Wiley & Sons, Inc., Hoboken, NJ.

2. Anslyn, E.V. and Dougherty, D.A. (2006) *Modern Physical Organic Chemistry*, University Science, Sausalito, CA.

3. Williams, A. (2003) *Free Energy Relationships in Organic and Bio-Organic Chemistry*, The Royal Society of Chemistry, Cambridge, UK.

4. Muller, P. (1994) Glossary of terms used in physical organic-chemistry. *Pure Appl. Chem.*, **66**, 1077–1184.

5. Hammond, G.S. (1955) A correlation of reaction rates. *J. Am. Chem. Soc.*, **77**, 334–338.

6. Leffler, J.E. (1955) The enthalpy-entropy relationship and its implications for organic chemistry. *J. Org. Chem.*, **20**, 1202–1231.

7. Leffler, J.E. (1955) Entropy requirements of the Hammett relationship. *J. Chem. Phys.*, **23**, 2199–2200.

8. Perez-Benito, J.F. (2013) Some tentative explanations for the enthalpy-entropy compensation effect in chemical kinetics: from experimental errors to the Hinshelwood-like model. *Monatsh. Chem.*, **144**, 49–58.

9. Eberson, L. (1982) Electron-transfer reactions in organic-chemistry. *Adv. Phys. Org. Chem.*, **18**, 79–185.

10. Albery, W.J. and Kreevoy, M.M. (1978) Methyl transfer reactions. *Adv. Phys. Org. Chem.*, **16**, 87–157.

11. Albery, W.J. (1979) Methyl transfer-reactions. *Pure Appl. Chem.*, **51**, 949–965.

12. Albery, W.J. (1980) The application of the Marcus relation to reactions in solution. *Annu. Rev. Phys. Chem.*, **31**, 227–263.

13. Dodd, J.A. and Brauman, J.I. (1984) On the application of the Marcus equation to methyl transfer (Sn2) reactions. *J. Am. Chem. Soc.*, **106**, 5356–5357.

14. Blowers, P. and Masel, R.I. (2000) Extensions of the Marcus equation for the prediction of approximate transition state geometries in hydrogen transfer and methyl transfer reactions. *Theor. Chem. Acc.*, **105**, 46–54.

15. Mayr, H., Breugst, M., and Ofial, A.R. (2011) Farewell to the HSAB treatment of ambident reactivity. *Angew. Chem. Int. Ed.*, **50**, 6470–6505.

16. Mayer, J.M. (2011) Understanding hydrogen atom transfer: from bond strengths to Marcus theory. *Acc. Chem. Res.*, **44**, 36–46.

17. Mayer, J.M. (2011) Simple Marcus-theory-type model for hydrogen-atom transfer/proton-coupled electron transfer. *J. Phys. Chem. Lett.*, **2**, 1481–1489.

18. Denisov, E.T. (2008) The Hammond postulate: a quantitative interpretation. *Russ. J. Phys. Chem. B*, **2**, 343–349.

19. Manz, T.A. and Sholl, D.S. (2010) A dimensionless reaction coordinate for quantifying the lateness of transition states. *J. Comput. Chem.*, **31**, 1528–1541.

20. Lewis, E.S. (1990) Rate-equilibrium LFER characterization of transition-states – the interpretation of alpha. *J. Phys. Org. Chem.*, **3**, 1–8.

21. Hammett, L.P. (1935) Some relations between reaction rates and equilibrium constants. *Chem. Rev.*, **17**, 125–136.

22. Burkhardt, G.N. (1935) Influence of substituents on organic reactions: a quantitative relationship. *Nature*, **136**, 684.

23. Shorter, J. (1995) The centenary of the birth of Louis Hammett. *Pure Appl. Chem.*, **67**, 835–840.

24. Shorter, J. (2000) The prehistory of the Hammett equation. *Chem. Listy*, **94**, 210–214.

25. Wells, P.R. (1963) Linear free energy relationships. *Chem. Rev.*, **63**, 171–219.

26. Jaffe, H.H. (1953) A reexamination of the Hammett equation. *Chem. Rev.*, **53**, 191–261.

27. Hammett, L.P. (1940) *Physical Organic Chemistry; Reaction Rates, Equilibria, and Mechanisms*, 1st edn, McGraw-Hill Book Company, Inc., New York, London.

28. Matsui, T., Ko, H.C., and Hepler, L.G. (1974) Thermodynamics of ionization of benzoic-acid and substituted benzoic-acids in relation to the Hammett equation. *Can. J. Chem.*, **52**, 2906–2911.

29. Eigen, M. (1965) Kinetics of proton transfer processes – general introduction. *Discuss. Faraday Soc.*, **39**, 7–15.

30. Eyring, E.M., Marshall, D.B., Strohbusch, F., and Suttinger, R. (1982) Proton-transfer reaction-rates and mechanisms. *ACS Sym. Ser.*, **198**, 63–80.

31. Mayr, H. and Ofial, A.R. (2006) The reactivity-selectivity principle: an imperishable myth in organic chemistry. *Angew. Chem. Int. Ed.*, **45**, 1844–1854.

32. Um, I.-H., Han, H.-J., Ahn, J.-A. *et al.* (2002) Reinterpretation of curved Hammett plots in reaction of nucleophiles with aryl benzoates: change in rate-determining step or mechanism versus ground-state stabilization. *J. Org. Chem.*, **67**, 8475–8480.

33. Santerre, G.M., Hansrote, C.J., and Crowell, T.I. (1958) The reaction of aromatic aldehydes with *n*-Butylamine. Acid catalysis and substituent effects. *J. Am. Chem. Soc.*, **80**, 1254–1257.

34. Kershaw, D.N. and Leisten, J.A. (1960) Alkyl-oxygen fission in ethyl carboxylates. *Proc. Chem. Soc.*, 84.

35. Hansch, C., Leo, A., and Taft, R.W. (1991) A survey of Hammett substituent constants and resonance and field parameters. *Chem. Rev.*, **91**, 165–195.

36. Hansch, C., Leo, A., and Hoekman, D.H. (1995) *Exploring QSAR*, American Chemical Society, Washington, DC.

37. Böhm, S., Fiedler, P., and Exner, O. (2004) Analysis of the ortho effect: acidity of 2-substituted benzoic acids. *New J. Chem.*, **28**, 67–74.

38. Böhm, S. and Exner, O. (2007) Substituent effects in the ortho position: model compounds with a removed reaction centre. *Pol. J. Chem.*, **81**, 993–1006.

39. Exner, O. and Böhm, S. (2006) Theory of substituent effects: recent advances. *Curr. Org. Chem.*, **10**, 763–778.

40. Rincon, L. and Almeida, R. (2012) Is the Hammett's constant free of steric effects? *J. Phys. Chem. A*, **116**, 7523–7530.

41. Jaffe, H.H. (1953) Some extensions of Hammett equation. *Science*, **118**, 246–247.

42. Kalfus, K., Kroupa, J., Vecera, M., and Exner, O. (1975) Additivity of substituent effects in meta-substituted and para-substituted benzoic-acids. *Collect. Czech. Chem. Commun.*, **40**, 3009–3019.

43. Wells, P.R., Ehrenson, S., and Taft, R.W. (1968) Substituent effects in the naphthalene series. An analysis of polar and pi delocalization effects. *Prog. Phys. Org. Chem.*, **6**, 147–322.

44. Parík, P. and Ludwig, M. (2004) Analysis of substituent effects in naphthalene skeleton. *Collect. Czech. Chem. Commun.*, **69**, 2297–2314.

45. Freitas, A.A., Dias, L.G., Macanita, A.A.L., and Quina, F.H. (2011) Substituent effects on the pH-dependent multiequilibria of flavylium salt analogs of anthocyanins. *J. Phys. Org. Chem.*, **24**, 1201–1208.

46. Vlasov, V.M. (2013) Varying the activation parameters as a method of determining whether there is an isokinetic relationship among the bimolecular nucleophilic reactions of benzene derivatives in solution. *Monatsh. Chem.*, **144**, 41–48.

47. Liu, L. and Guo, Q.X. (2001) Isokinetic relationship, isoequilibrium relationship, and enthalpy-entropy compensation. *Chem. Rev.*, **101**, 673–695.

48. Exner, O. (1964) Concerning the isokinetic relationship. *Nature*, **201**, 488–490.

49. Exner, O. (1970) Determination of the isokinetic temperature. *Nature*, **227**, 366–367.

50. Exner, O. (1973) Conception of reactivity and meaning of isokinetic relation. *Chem. Listy*, **67**, 135–158.

51. Hepler, L.G. (1971) Thermodynamic analysis of Hammett equation, the temperature dependence of rho, and the isoequilibrium (isokinetic) relationship. *Can. J. Chem.*, **49**, 2803–2807.

52. Ouvrard, C., Berthelot, M., Lamer, T., and Exner, O. (2001) A program for linear regression with a common point of intersection: the isokinetic relationship. *J. Chem. Inf. Comput. Sci.*, **41**, 1141–1144.

53. Vlasov, V.M. and Kornakova, T.A. (2013) Isokinetic relationships for SNAr reactions in different solvents. *J. Phys. Org. Chem.*, **26**, 131–136.

54. Linert, W. and Jameson, R.F. (1989) The isokinetic relationship. *Chem. Soc. Rev.*, **18**, 477–505.

55. Linert, W. (1994) Mechanistic and structural investigations based on the isokinetic relationship. *Chem. Soc. Rev.*, **23**, 429–438.

56. Bitterwolf, T.E., Linder, R.E., and Ling, A.C. (1970) Temperature dependence of the Hammett reaction constant. *J. Chem. Soc. B*, 1673–1674.

57. Swain, C.G. and Lupton, E.C. (1968) Field and resonance components of substituent effects. *J. Am. Chem. Soc.*, **90** (16), 4328–4337.

58. Jiang, X.-K. (1997) Establishment and successful application of the $\sigma_{JJ}\cdot$ scale of spin-delocalization substituent constants. *Acc. Chem. Res.*, **30**, 283–289.

59. Okamoto, Y. and Brown, H.C. (1957) A quantitative treatment for electrophilic reactions of aromatic derivatives. *J. Org. Chem.*, **22**, 485–494.

60. Oehlke, A., Auer, A.A., Schreiter, K. *et al.* (2009) Electrophilic substituent constant σ^+ of electron donor substituents, in nonpolar media. *J. Org. Chem.*, **74**, 3316–3322.

61. Stock, L.M. and Brown, H.C. (1963) A quantitative treatment of directive effects in aromatic substitution. *Adv. Phys. Org. Chem.*, **1**, 35–154.

62. Koshy, K.M., Roy, D., and Tidwell, T.T. (1979) Substituent effects of the trifluoromethyl group on electrophilic additions to alkenes – solvolysis of the trifluoromethyl group – protonation of alkenes less basic than ethylene, P+ values of deactivated styrenes, and reactivity-selectivity effects. *J. Am. Chem. Soc.*, **101**, 357–363.

63. Kim, C.J. and Brown, H.C. (1969) Rates and products of acetolysis of threo-3-aryl-2-butyl bro-sylates. An experimental approach to the magnitude of rate accelerations attributable to aryl participation in the acetolysis and the stereochemical consequences. *J. Am. Chem. Soc.*, **91**, 4289–4291.

64. Yukawa, Y. and Tsuno, Y. (1959) Resonance effect in Hammett relationship .3. The modified Hammett relationship for electrophilic reactions. *Bull. Chem. Soc. Jpn.*, **32**, 971–981.

65. Nakata, K., Fujio, M., Nishimoto, K., and Tsuno, Y. (2003) Theoretical studies on empirical structure-reactivity relationship: the Yukawa-Tsuno equation. *J. Phys. Org. Chem.*, **16**, 323–335.

66. Nakata, K., Fujio, M., Nishimoto, K., and Tsuno, Y. (2013) Theoretical study of substituent effects on the gas-phase stabilities of phenoxide anions. *J. Phys. Org. Chem.*, **26**, 115–123.

67. Taft, R.W. (1952) Polar and steric substituent constants for aliphatic and O-Benzoate groups from rates of esterification and hydrolysis of esters. *J. Am. Chem. Soc.*, **74**, 3120–3128.

68. Taft, R.W. (1953) Linear steric energy relationships. *J. Am. Chem. Soc.*, **75**, 4538–4539.

69. Roberts, J.D. and Moreland, W.T. (1953) Electrical effects of substituent groups in saturated systems – reactivities of 4-substituted bicyclo [2.2.2]octane-1-carboxylic acids. *J. Am. Chem. Soc.*, **75**, 2167–2173.

70. Grob, C.A. (1985) Inductive charge dispersal in quinuclidinium ions .13. Polar effects. *Helv. Chim. Acta*, **68**, 882–886.

71. Bielmann, R., Grob, C.A., Kury, D., and Yao, G.W. (1985) Inductivity and charge dispersal in quinuclidinium and bicyclo[2.2.2]octyl-1-cations. *Tetrahedron Lett.*, **26**, 315–318.

72. Charton, M. (1981) Electrical effect substituent constants for correlation analysis. *Prog. Phys. Org. Chem.*, **13**, 119–251.

73. Exner, O. (1999) The inductive effect: theory and quantitative assessment. *J. Phys. Org. Chem.*, **12**, 265–274.

74. Swain, C.G., Unger, S.H., Rosenquist, N.R., and Swain, M.S. (1983) Substituent effects on chemical-reactivity – improved evaluation of field and resonance components. *J. Am. Chem. Soc.*, **105**, 492–502.

75. Cao, C.Z. and Wu, Y.X. (2013) Recent progress in quantifying substituent effects. *Sci. China Chem.*, **56**, 883–910.

76. Cherkasov, A. (2005) 'Inductive' descriptors: 10 successful years in QSAR. *Curr. Comput. Aided Drug Des.*, **1**, 21–42.

77. Rezende, M.C. and Millán, D. (2011) Theoretical substituent electrophilicities. *J. Braz. Chem. Soc.*, **22** (11), 2078–2086.

78. Segala, M., Takahata, Y., and Chong, D.P. (2006) Geometry, solvent, and polar effects on the relationship between calculated core-electron binding energy shifts (ΔCEBE) and Hammett substituent (σ) constants. *J. Mol. Struc. Theochem.*, **758**, 61–69.

79. Lindberg, B., Svensson, S., Malmquist, P.A. *et al.* (1976) Correlation of ESCA shifts and Hammett substituent constants in substituted benzene-derivatives. *Chem. Phys. Lett.*, **40**, 175–179.

80. Takahata, Y. (2008) Inductive and resonance effects based on core-electron binding energy shift. *Int. J. Quantum Chem.*, **108**, 2326–2333.

81. Takahata, Y. and Marques, A.D. (2010) Accurate core-electron binding energy shifts from density functional theory. *J. Electron. Spectrosc. Relat. Phenom.*, **178**, 80–87.

82. Wang, C., Fu, Y., and Liu, L. (2010) Theoretical estimation of Hammett sigma (p) constants of organic radical groups. *Chinese Sci. Bull.*, **55**, 2904–2908.

83. Afanas'ev, I.B. (1971) Correlation equations in free-radical reactions. *Russ. Chem. Rev.*, **40**, 216–232.

84. Modglin, J.D., Erdely, V.K., Lin, C.Y. *et al.* (2011) Hammett correlations in the chemistry of 3-phenylpropyl radicals. *J. Phys. Chem. A*, **115**, 14687–14696.

85. Yoshida, T., Hirozumi, K., Harada, M. *et al.* (2011) Density functional theory study of hydrogen atom abstraction from a series of para-substituted phenols: why is the Hammett σ_p^+ constant able to represent radical reaction rates? *J. Org. Chem.*, **76**, 4564–4570.

86. Zhang, Y.L., Wang, L., Moss, R.A., and Platz, M.S. (2009) Ultrafast spectroscopy of arylchlorodiazirines: Hammett correlations of excited state lifetimes. *J. Am. Chem. Soc.*, **131**, 16652–16653.

87. Sadlej-Sosnowska, N. and Kijak, M. (2012) Excited state substituent constants: to Hammett or not? *Struct. Chem.*, **23**, 359–365.

88. Ciscato, L.F.M.L., Bartoloni, F.H., Weiss, D. *et al.* (2010) Experimental evidence of the occurrence of intramolecular electron transfer in catalyzed 1,2-dioxetane decomposition. *J. Org. Chem.*, **75**, 6574–6580.

89. Lewis, M., Bagwill, C., and Hardebeck, L.K.E. (2012) The use of Hammett constants to understand the non-covalent binding of aromatics. *Comput. Struct. Biotechnol. J.*, **1**, e201204004.

90. Harper, K.C. and Sigman, M.S. (2013) Using physical organic parameters to correlate asymmetric catalyst performance. *J. Org. Chem.*, **78**, 2813–2818.

91. Gottardo, C., Kraft, T.M., Hossain, M.S. *et al.* (2008) Linear free-energy correlation analysis of the electronic effects of the substituents in the Sonogashira coupling reaction. *Can. J. Chem.*, **86**, 410–415.

92. Rao, B.V., Kwon, K.Y., Liu, A.W., and Bartels, L. (2004) Measurement of a linear free energy relationship one molecule at a time. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 17920–17923.

93. Selassie, C.D., Mekapati, S.B., and Verma, R.P. (2002) QSAR: then and now. *Curr. Top. Med. Chem.*, **2**, 1357–1379.

94. Verloop, A. (1987) *The STERIMOL Approach to Drug Design*, Marcel Dekker, New York, NY.

95. STERIMOL (2013) *Quantum Chemistry Program Exchange*, http://www.ccl.net/cca/software/ SOURCES/FORTRAN/STERIMOL/ (accessed 31 March 2016).

96. Harper, K.C., Bess, E.N., and Sigman, M.S. (2012) Multidimensional steric parameters in the analysis of asymmetric catalytic reactions. *Nat. Chem.*, **4**, 366–374.

97. Mayr, H. and Ofial, A.R. (2008) Do general nucleophilicity scales exist? *J. Phys. Org. Chem.*, **21**, 584–595.

98. Swain, C.G. and Scott, C.B. (1953) Concerted displacement reactions .10. Quantitative correlation of relative rates – comparison of hydroxide ion with other nucleophilic reagents toward alkyl halides, esters, epoxides and acyl halides. *J. Am. Chem. Soc.*, **75**, 141–147.

99. Ritchie, C.D. (1975) Cation-anion combination reactions .13. correlation of reactions of nucleophiles with esters. *J. Am. Chem. Soc.*, **97**, 1170–1179.

100. Ritchie, C.D. (1972) Nucleophilic reactivities toward cations. *Acc. Chem. Res.*, **5**, 348–354.

101. Mayr, H. (2011) Reply to T. W. Bentley: limitations of the s(E+N) and related equations. *Angew. Chem. Int. Ed.*, **50**, 3612–3618.

102. Mayr, H. and Patz, M. (1994) Scales of nucleophilicity and electrophilicity – a system for ordering polar organic and organometallic reactions. *Angew. Chem. Int. Ed.*, **33**, 938–957.

103. Mayr, H. *Database of Nucleophilicities and Electrophilicities*. http://www.cup.uni-muenchen .de/oc/mayr/DBintro.html (accessed 5 October 2013).

104. Zhuo, L.-G., Lian, W., and Yu, Z.-X. (2012) A frontier molecular orbital theory approach to understanding the Mayr equation and to quantifying nucleophilicity and electrophilicity by using HOMO and LUMO energies. *Asian J. Org. Chem.*, **1**, 336–345.

105. Chamorro, E., Duque-Norena, M., Notario, R., and Perez, P. (2013) Intrinsic relative scales of electrophilicity and nucleophilicity. *J. Phys. Chem. A*, **117**, 2636–2643.

106. Chattaraj, P.K., Giri, S., and Duley, S. (2011) Update 2 of: electrophilicity index. *Chem. Rev.*, **111**, Pr43–Pr75.

107. Rezende, M.C., Dominguez, M., Aracena, A., and Millan, D. (2011) Solvatochromism and electrophilicity. *Chem. Phys. Lett.*, **514**, 267–273.

108. Rezende, M.C. and Aracena, A. (2012) Electrophilicity and solvatochromic reversal of pyridinium phenolate betaine dyes. *Chem. Phys. Lett.*, **542**, 147–152.

109. Rezende, M.C. and Aracena, A. (2013) A general framework for the solvatochromism of pyridinium phenolate betaine dyes. *Chem. Phys. Lett.*, **558**, 77–81.

110. Reichardt, C. and Welton, T. (2011) *Solvents and Solvent Effects in Organic Chemistry*, 4th edn, Wiley-VCH, Weinheim, Germany.

111. Abboud, J.-L.M. and Notario, R. (1999) Critical compilation of scales of solvent parameters. Part I. Pure, non-hydrogen bond donor solvents. *Pure Appl. Chem.*, **71**, 645–718.

112. Stenutz, R. *Tables for Organic Chemistry: Polarity Scales*, http://www.stenutz.eu/chem/ (accessed 5 October 2013).

113. Kamlet, M.J., Abboud, J.L.M., Abraham, M.H., and Taft, R.W. (1983) Linear solvation energy relationships .23. A comprehensive collection of the solvatochromic parameters, π^*, α and β, and some methods for simplifying the generalized solvatochromic equation. *J. Org. Chem.*, **48**, 2877–2887.

114. Catalán, R. *Web Solvents*, http://www.uam.es/personal_pdi/ciencias/catalan/web_solvents.htm (accessed 22 August 2013).

115. Marcus, Y. (1993) The properties of organic liquids that are relevant to their use as solvating solvents. *Chem. Soc. Rev.*, **22**, 409–416.

116. Malavolta, L., Pinto, M.R.S., Cuvero, J.H., and Nakaie, C.R. (2006) Interpretation of the dissolution of insoluble peptide sequences based on the acid–base properties of the solvent. *Protein Sci.*, **15**, 1476–1488.

117. Swain, C.G., Swain, M.S., Powell, A.L., and Alunni, S. (1983) Solvent effects on chemical-reactivity – evaluation of anion and cation solvation components. *J. Am. Chem. Soc.*, **105**, 502–513.

118. Rezende, M.C. and Aracena, A. (2012) In search of the thermo/halochromism of the $E_T(30)$ pyridinium-N-phenolate betaine dye. *Spectrochim. Acta A Mol. Biomol. Spectrosc.*, **98**, 18–22.

119. Fidale, L.C., Heinze, T., and El Seoud, O.A. (2013) Perichromism: a powerful tool for probing the properties of cellulose and its derivatives. *Carbohydr. Polym.*, **93**, 129–134.

120. Rezende, M.C., Dominguez, M., and Aracena, A. (2012) The positive halochromism of phenolate dyes in hydroxylic solutions of tetraalkylammonium cations. *Spectrochim. Acta A Mol. Biomol. Spectrosc.*, **87**, 61–66.

121. Bastos, E.L., Bartoloni, F.H., and Gonçalves, L.C.P. (2013) Acid–base and solvation properties of metal phenolates, in *PATAI'S Chemistry of Functional Groups* (ed J. Zabicky), John Wiley & Sons, Ltd, Chichester, UK, pp. 1–72.

122. Scherrer, R.A. and Howard, S.M. (1977) Use of distribution coefficients in quantitative structure-activity-relationships. *J. Med. Chem.*, **20**, 53–58.

123. Fujita, T., Hansch, C., and Iwasa, J. (1964) New substituent constant, π, derived from partition coefficients. *J. Am. Chem. Soc.*, **86**, 5175–5180.

124. OECD (2004) *Test No. 117: Partition Coefficient (n-octanol/water), HPLC Method*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264069824-en (accessed 5 October 2013).

125. Mannhold, R., Poda, G.I., Ostermann, C., and Tetko, I.V. (2009) Calculation of molecular lipophilicity: state-of-the-art and comparison of log P methods on more than 96,000 compounds. *J. Pharm. Sci.*, **98**, 861–893.

126. Mannhold, R. and van de Waterbeemd, H. (2001) Substructure and whole molecule approaches for calculating log P. *J. Comput. Aided Mol. Des.*, **15**, 337–354.

127. Tetko, I.V., Poda, G.I., Ostermann, C., and Mannhold, R. (2009) Large-scale evaluation of log P predictors: local corrections may compensate insufficient accuracy and need of experimentally testing every other compound. *Chem. Biodivers.*, **6**, 1837–1844.

128. Syracuse Research Corporation (SRC) *PHYSPROP Database,* http://srcinc.com/what-we-do/free-demos.aspx (accessed 22 September 2013).

129. ALOGPS. *Virtual Computational Chemistry Laboratory*, http://www.vcclab.org/ (accessed 22 September 2013).

130. Tetko, I.V., Gasteiger, J., Todeschini, R. *et al.* (2005) Virtual computational chemistry laboratory – design and description. *J. Comput. Aided Mol. Des.*, **19**, 453–463.

131. U.S. Environmental Protection Agency. *EPA Estimation Program Interface (EPI)*Suite: https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface (accessed 31 March 2016).

132. Martel, S., Gillerat, F., Carosati, E. *et al.* (2013) Large, chemically diverse dataset of logP measurements for benchmarking studies. *Eur. J. Pharm. Sci.*, **48**, 21–29.

133. Zviling, M., Leonov, H., and Arkin, I.T. (2005) Genetic algorithm-based optimization of hydrophobicity tables. *Bioinformatics*, **21**, 2651–2656.

134. White, S.H. and Wimley, W.C. (1999) Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.*, **28**, 319–365.

135. White, S. H. *Wimley-White Hydropathy Scales,* http://blanco.biomol.uci.edu/hydrophobicity_scales.html (accessed 28 August 2013).

136. White, S. H. *Membrane Protein Explorer (MPEx)*, http://blanco.biomol.uci.edu/mpex/index.html (accessed 28 August 2013).

137. Abraham, M.H., Grellier, P.L., Abboud, J.L.M. *et al.* (1988) Solvent effects in organic-chemistry – recent developments. *Can. J. Chem.*, **66**, 2673–2686.

138. Jessop, P.G., Jessop, D.A., Fu, D.B., and Phan, L. (2012) Solvatochromic parameters for solvents of interest in green chemistry. *Green Chem.*, **14**, 1245–1259.

139. Lee, J.M., Ruckes, S., and Prausnitz, J.M. (2008) Solvent polarities and Kamlet-Taft parameters for ionic liquids containing a pyridinium cation. *J. Phys. Chem. B*, **112**, 1473–1476.

140. Lu, J., Brown, J.S., Liotta, C.L., and Eckert, C.A. (2001) Polarity and hydrogen-bonding of ambient to near-critical water: Kamlet-Taft solvent parameters. *Chem. Commun.*, 665–666.

141. Minami, K., Mizuta, M., Suzuki, M. *et al.* (2006) Determination of Kamlet-Taft solvent parameters π^* of high pressure and supercritical water by the UV–vis absorption spectral shift of 4-nitroanisole. *Phys. Chem. Chem. Phys.*, **8**, 2257–2264.

142. Abbott, A.P. and Eardley, C.A. (1998) Solvent properties of liquid and supercritical 1,1,1,2-tetrafluoroethane. *J. Phys. Chem. B*, **102**, 8574–8578.

143. Abbott, A.P., Hope, E.G., and Palmer, D.J. (2007) Probing solute clustering in supercritical solutions using solvatochromic parameters. *J. Phys. Chem. B*, **111**, 8119–8125.

144. Lu, J., Brown, J.S., Boughner, E.C. *et al.* (2002) Solvatochromic characterization of near-critical water as a benign reaction medium. *Ind. Eng. Chem. Res.*, **41**, 2835–2841.

145. Kruse, A. and Dinjus, E. (2007) Hot compressed water as reaction medium and reactant – properties and synthesis reactions. *J. Supercrit. Fluid*, **39**, 362–380.

146. Townsend, S.H., Abraham, M.A., Huppert, G.L. *et al.* (1988) Solvent effects during reactions in supercritical water. *Ind. Eng. Chem. Res.*, **27**, 143–149.

147. Abraham, M.H., Doherty, R.M., Kamlet, M.J. *et al.* (1987) Linear solvation energy relationships .37. An analysis of contributions of dipolarity polarizability, nucleophilic assistance, electrophilic assistance, and cavity terms to solvent effects on tert-butyl halide solvolysis rates. *J. Chem. Soc. Perk. Trans.*, **2**, 913–920.

148. El Seoud, O.A. (2010) Solvation simplified. *Quim. Nova*, **33** (10), 2187–2192.

149. El Seoud, O.A. (2009) Understanding solvation. *Pure Appl. Chem.*, **81**, 697–707.

150. Sato, B.M., de Oliveira, C.G., Martins, C.T., and El Seoud, O.A. (2010) Thermo-solvatochromism in binary mixtures of water and ionic liquids: on the relative importance of solvophobic interactions. *Phys. Chem. Chem. Phys.*, **12**, 1764–1771.

151. Sato, B.M., Martins, C.T., and El Seoud, O.A. (2012) Solvation in aqueous binary mixtures: consequences of the hydrophobic character of the ionic liquids and the solvatochromic probes. *New J. Chem.*, **36**, 2353–2360.

152. Silva, P.L., Bastos, E.L., and El Seoud, O.A. (2007) Solvation in binary mixtures of water and polar aprotic solvents: theoretical calculations of the concentrations of solvent-water hydrogen-bonded species and application to thermosolvatochromism of polarity probes. *J. Phys. Chem. B*, **111**, 6173–6180; **112**, 187.

153. Silva, P.L., Trassi, M.A.S., Martins, C.T., and El Seoud, O.A. (2009) Solvatochromism in binary mixtures: first report on a solvation free energy relationship between solvent exchange equilibrium constants and the properties of the medium. *J. Phys. Chem. B*, **113**, 9512–9519.

154. Bastos, E.L., Silva, P.L., and El Seoud, O.A. (2006) Thermosolvatochromism of betaine dyes revisited: theoretical calculations of the concentrations of alcohol−water hydrogen-bonded species and application to solvation in aqueous alcohols. *J. Phys. Chem. A*, **110**, 10287–10295.

155. Martins, C.T., Lima, M.S., Bastos, E.L., and El Seoud, O.A. (2008) Thermo-solvatochromism of merocyanine polarity probes – what are the consequences of increasing probe lipophilicity through annelation? *Eur. J. Org. Chem.*, **2008**, 1165–1180.

156. Catalán, R. (2001) Solvent effects based on pure solvent scales, in *Handbook of Solvents* (ed G. Wypych), ChemTec, Toronto; New York, pp. 583–616.

157. Kevill, D.N. and D'Souza, M.J. (2008) Sixty years of the Grunwald-Winstein equation: development and recent applications. *J. Chem. Res. S*, 61–66.

158. Kevill, D.N. and D'Souza, M.J. (2010) Use of the simple and extended Grunwald-Winstein equations in the correlation of the rates of solvolysis of highly hindered tertiary alkyl derivatives. *Curr. Org. Chem.*, **14**, 1037–1049.

159. Streidl, N. and Mayr, H. (2011) Ionizing power of aprotic solvents. *Eur. J. Org. Chem.*, **2011**, 2498–2506.

160. Kevill, D.N. and Anderson, S.W. (1991) An improved scale of solvent nucleophilicity based on the solvolysis of the S-methyldibenzothiophenium ion. *J. Org. Chem.*, **56**, 1845–1850.

161. Marcus, Y. (1984) The effectivity of solvents as electron pair donors. *J. Solution Chem.*, **13**, 599–624.

162. Schmeisser, M., Illner, P., Puchta, R. *et al.* (2012) Gutmann donor and acceptor numbers for ionic liquids. *Chem. Eur. J.*, **18**, 10969–10982.

163. Linert, W., Jameson, R.F., and Taha, A. (1993) Donor numbers of anions in solution – the use of solvatochromic Lewis-acid base indicators. *J. Chem. Soc. Dalton*, 3181–3186.

164. Linert, W. and Jameson, R.F. (1993) Acceptor properties of solvents – the use of isokinetic relationships to elucidate the relationship between the acceptor number and the solvatochromism of N-phenolate betaine dyes. *J. Chem. Soc. Perk. Trans.*, **2**, 1415–1421.

165. Linert, W., Jameson, R.F., Bauer, G., and Taha, A. (1997) Estimation of the acceptor numbers of cations by means of an acid–base indicator. *J. Coord. Chem.*, **42**, 211–229.

166. Swain, C.G. (1984) Substituent and solvent effects on chemical-reactivity. *J. Org. Chem.*, **49**, 2005–2010.

167. Abraham, M.H. (1993) Scales of solute hydrogen-bonding – their construction and application to physicochemical and biochemical processes. *Chem. Soc. Rev.*, **22**, 73–83.

168. Stephens, T.W., Willis, B., Dabadge, N. *et al.* (2013) Corrrelations of solute partititoning and enthalpies of solvation for organic solutes in ionic liquids using a temperature independent free energy relationship. *Eur. Chem. Bull.*, **2**, 887–897.

169. Poole, C.F., Atapattu, S.N., Poole, S.K., and Bell, A.K. (2009) Determination of solute descriptors by chromatographic methods. *Anal. Chim. Acta*, **652**, 32–53.

170. Abraham, M.H. and McGowan, J.C. (1987) The use of characteristic volumes to measure cavity terms in reversed phase liquid-chromatography. *Chromatographia*, **23**, 243–246.

171. Hoffmann, E.A., Rajko, R., Fekete, Z.A., and Kortvelyesi, T. (2009) Quantum chemical characterization of Abraham solvation parameters for gas–liquid chromatographic stationary phases. *J. Chromatogr. A*, **1216**, 8535–8544.

172. Endo, S., Watanabe, N., Ulrich, N., *et al.* (2015) UFZ-LSER database v 2.1 [Internet], Leipzig, Germany, Helmholtz Centre for Environmental Research-UFZ, https://www.ufz.de/index.php?en=31698&contentonly=1&lserd_data[mvc]=Public/start (accessed 31 March 2016).

173. Palatinus, J.A., Carroll, F.A., Argenton, A.B., and Quina, F.H. (2006) An improved characteristic molecular volume parameter for linear solvation energy relationships of acyclic alkanes. *J. Phys. Org. Chem.*, **19**, 725–730.

174. Mintz, C. and Acree, W.E. (2007) Comments on 'an improved characteristic molecular volume parameter for linear solvation energy relationships of acyclic alkanes'. *J. Phys. Org. Chem.*, **20**, 365–367.

175. van Noort, P.C.M., Haftka, J.J.H., and Parsons, J.R. (2011) A simple McGowan specific volume correction for branching in hydrocarbons and its consequences for some other solvation parameter values. *Chemosphere*, **84**, 1102–1107.

176. van Noort, P.C.M. (2012) Solvation thermodynamics and the physical-chemical meaning of the constant in Abraham solvation equations. *Chemosphere*, **87**, 125–131.

177. Abraham, M.H. and Acree, W.E. (2012) The hydrogen bond properties of water from 273 K to 573 K; equations for the prediction of gas-water partition coefficients. *Phys. Chem. Chem. Phys.*, **14**, 7433–7440.

178. Lagalante, A.F., Clarke, A.M., and Bruno, T.J. (1998) Modeling the water-R134a partition coefficients of organic solutes using a linear solvation energy relationship. *J. Phys. Chem. B*, **102**, 8889–8892.

179. Quina, F.H., Alonso, E.O., and Farah, J.P.S. (1995) Incorporation of nonionic solutes into aqueous micelles – a linear solvation free-energy relationship analysis. *J. Phys. Chem.*, **99**, 11708–11714.

180. Sprunger, L., Acree, W.E., and Abraham, M.H. (2007) Linear free energy relationship correlation of the distribution of solutes between water and sodium dodecyl sulfate (SDS) micelles and between gas and SDS micelles. *J. Chem. Inf. Model.*, **47**, 1808–1817.

181. Abraham, M.H., Andonianhaftvan, J., Whiting, G.S. *et al.* (1994) Hydrogen-bonding .34. The factors that influence the solubility of gases and vapors in water at 298-K, and a new method for its determination. *J. Chem. Soc. Perk. Trans.*, **2**, 1777–1791.

182. Freitas, A.A., Quina, F.H., and Carroll, F.A. (2000) A linear free energy analysis of the surface tension of organic liquids. *Langmuir*, **16**, 6689–6692.

183. Freitas, A.A., Quina, F.H., and Carroll, F.A. (1997) Estimation of water-organic interfacial tensions. A linear free energy relationship analysis of interfacial adhesion. *J. Phys. Chem. B*, **101**, 7488–7493.

184. Tomlinson-Phillips, J., Davis, J., Ben-Amotz, D. *et al.* (2011) Structure and dynamics of water dangling OH bonds in hydrophobic hydration shells. Comparison of simulation and experiment. *J. Phys. Chem. A*, **115**, 6177–6183.

185. Hsieh, C.S., Campen, R.K., Verde, A.C.V. *et al.* (2011) Ultrafast reorientation of dangling OH groups at the air-water interface using femtosecond vibrational spectroscopy. *Phys. Rev. Lett.*, **107**, 116102.

186. Perera, P.N., Fega, K.R., Lawrence, C. *et al.* (2009) Observation of water dangling OH bonds around dissolved nonpolar groups. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 12230–12234.

187. Green, A.J., Perry, A., Moore, P.B., and Space, B. (2012) A theoretical study of the sum frequency vibrational spectroscopy of the carbon tetrachloride/water interface. *J. Phys. Condens. Matter*, **24**, 124108.

188. Shen, Y.R. and Ostroverkhov, V. (2006) Sum-frequency vibrational spectroscopy on water interfaces: polar orientation of water molecules at interfaces. *Chem. Rev.*, **106**, 1140–1154.

189. Sprunger, L., Acree, W.E., and Abraham, M.H. (2007) Comment on systematic investigation of the sorption properties of polyurethane foams for organic vapors. *Anal. Chem.*, **79**, 6891–6893.

190. Mutelet, F., Ortega-Villa, V., Moise, J.C. *et al.* (2011) Prediction of partition coefficients of organic compounds in ionic liquids using a temperature-dependent linear solvation energy relationship with parameters calculated through a group contribution method. *J. Chem. Eng. Data*, **56**, 3598–3606.

191. Mintz, C., Ladlie, T., Burton, K. *et al.* (2008) Characterization of the partitioning of gaseous solutes into humic acid with the Abraham model and temperature-independent equation coefficients. *QSAR Comb. Sci.*, **27**, 483–491.

192. Masson, E., Ling, X.X., Joseph, R. *et al.* (2012) Cucurbituril chemistry: a tale of supramolecular success. *RSC Adv.*, **2**, 1213–1247.

193. Ghasemi, J.B., Salahinejad, M., and Rofouei, M.K. (2011) Review of the quantitative structure-activity relationship modelling methods on estimation of formation constants of macrocyclic compounds with different guest molecules. *Supramol. Chem.*, **23**, 614–629.

194. Pérez-Garrido, A., Helguera, A.M., Guillen, A.A. *et al.* (2009) Convenient QSAR model for predicting the complexation of structurally diverse compounds with beta-cyclodextrins. *Bioorg. Med. Chem.*, **17**, 896–904.

195. Pérez-Garrido, A., Helguera, A.M., Cordeiro, M.N.D.S., and Escudero, A.G. (2009) QSPR modelling with the topological substructural molecular design approach: beta-cyclodextrin complexation. *J. Pharm. Sci.*, **98**, 4557–4576.

196. Blanford, W.J., Gao, H., Dutta, M., and Ledesma, E.B. (2013) Solubility enhancement and QSPR correlations for polycyclic aromatic hydrocarbons complexation with α, β, and γ cyclodextrins. *J. Incl. Phenom. Macrocycl. Chem.*, **78**, 415–427, 10.1007/s10847-013-0313-2.

197. Gao, H., Blanford, W.J., and Gao, A.F. (2013) Solubility enhancement effect of cyclodexin on groundwater pollutants. *Ground Water*, **51**, 268–275.

198. Sang, P., Zou, J.-W., Dai, D.-M. *et al.* (2013) Prediction of the complexation of structurally diverse compounds with β-cyclodextrin using structural descriptors derived from electrostatic potentials on molecular surface and different chemometric methods. *Chemometr. Intell. Lab. Syst.*, **127**, 166–176.

199. Ashton, P.R., Fyfe, M.C.T., Hickingbottom, S.K. *et al.* (1998) Hammett correlations 'beyond the molecule'. *J. Chem. Soc. Perk. Trans.*, **2**, 2117–2128.

200. Smithrud, D.B. and Diederich, F. (1990) Strength of molecular complexation of apolar solutes in water and in organic-solvents is predictable by linear free-energy relationships – a general-model for solvation effects on apolar binding. *J. Am. Chem. Soc.*, **112**, 339–343.

201. Smithrud, D.B., Sanford, E.M., Chao, I. *et al.* (1990) Solvent effects in molecular recognition. *Pure Appl. Chem.*, **62**, 2227–2236.

202. D'Souza, V.T. and Bender, M.L. (1987) Miniature organic-models of enzymes. *Acc. Chem. Res.*, **20**, 146–152.

203. Raj, V., Chandrakala, T., and Rajasekaran, K. (2008) Guest-host interactions in the cleavage of phenylphenyl acetates by beta-cyclodextrin in alkaline medium. *J. Chem. Sci.*, **120**, 325–328.

204. Basilio, N., Garcia-Rio, L., Moreira, J.A., and Pessêgo, M. (2010) Supramolecular catalysis by Cucurbit[7]uril and cyclodextrins: similarity and differences. *J. Org. Chem.*, **75**, 848–855.

205. Verma, J., Khedkar, V.M., and Coutinho, E.C. (2010) 3D-QSAR in drug design – a review. *Curr. Top. Med. Chem.*, **10**, 95–115.

206. Free, S.M. and Wilson, J.W. (1964) Mathematical contribution to structure-activity studies. *J. Med. Chem.*, **7**, 395–399.

207. Kubinyi, H. (1997) QSAR and 3D QSAR in drug design .1. Methodology. *Drug Discov. Today*, **2**, 457–467.

208. Chen, H.M., Carlsson, L., Eriksson, M. *et al.* (2013) Beyond the scope of Free-Wilson analysis: building interpretable QSAR models with machine learning algorithms. *J. Chem. Inf. Model.*, **53**, 1324–1336.

209. Hansch, C. and Fujita, T. (1964) ρ-σ-π analysis. Method for correlation of biological activity + chemical structure. *J. Am. Chem. Soc.*, **86**, 1616–1626.

210. Hansch, C. and Clayton, J.M. (1973) Lipophilic character and biological-activity of drugs .2. Parabolic case. *J. Pharm. Sci.*, **62**, 1–21.

211. Kubinyi, H. (1976) Quantitative structure-activity-relationships .2. Mixed approach, based on Hansch and Free-Wilson analysis. *J. Med. Chem.*, **19**, 587–600.

212. Abraham, M.H. (2011) The permeation of neutral molecules, ions, and ionic species through membranes: brain permeation as an example. *J. Pharm. Sci.*, **100**, 1690–1701.

213. Abraham, M.H., Ibrahim, A., and Acree, W.E. (2006) Air to brain, blood to brain and plasma to brain distribution of volatile organic compounds: linear free energy analyses. *Eur. J. Med. Chem.*, **41**, 494–502.

214. Abraham, M.H., Sanchez-Moreno, R., Cometto-Muniz, J.E., and Cain, W.S. (2007) A quantitative structure-activity analysis on the relative sensitivity of the olfactory and the nasal trigeminal chemosensory systems. *Chem. Senses*, **32**, 711–719.

215. Abraham, M.H., Sanchez-Moreno, R., Cometto-Muniz, J.E., and Cain, W.S. (2012) An algorithm for 353 odor detection thresholds in humans. *Chem. Senses*, **37**, 207–218.

216. Hoover, K.R., Flanagan, K.S., Acree, W.E., and Abraham, M.H. (2007) Chemical toxicity correlations for several protozoas, bacteria, and water fleas based on the Abraham solvation parameter model. *J. Environ. Eng. Sci.*, **6**, 165–174.

217. Ballabio, D., Manganaro, A., Consonni, V. *et al.* (2009) Introduction to MOLE DB – on-line molecular descriptors database. *Match-Commun. Math. Co.*, **62**, 199–207.

218. Consonni, V. and Todeschini, R. (2000) *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim; New York.

219. Sushko, I., Novotarskyi, S., Korner, R. *et al.* (2011) Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. *J. Comput. Aided Mol. Des.*, **25**, 533–554.

220. Todeschini, R., Consonni, V., and Gramatica, P. (2009) Chemometrics in QSAR, in *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, vol. **4** (eds S. Brown, R. Tauler, and R. Walczak), Elsevier, Amsterdam, pp. 129–172.

221. Berhanu, W.M., Pillai, G.G., Oliferenko, A.A., and Katritzky, A.R. (2012) Quantitative structure-activity/property relationships: the ubiquitous links between cause and effect. *Chempluschem*, **77**, 507–517.

222. Katritzky, A.R., Kuanar, M., Slavov, S. *et al.* (2010) Quantitative correlation of physical and chemical properties with chemical structure: utility for prediction. *Chem. Rev.*, **110**, 5714–5789.

223. *Molecular Descriptors Website*, http://www.moleculardescriptors.eu/index.htm (accessed 22 August 2013).

224. *Milano Chemometrics and QSAR Research Group. Molecular Descriptors Data Base (MOLE db)*, http://www.moleculardescriptors.eu/index.htm (accessed 22 August 2013).

225. Talete. *Dragon 6 Software for Molecular Descriptor Calculation*, http://www.talete.mi.it/products/dragon_description.htm (accessed 5 October 2013).

226. Gramatica, P., Cassani, S., Roy, P.P. *et al.* (2012) QSAR modeling is not "Push a Button and Find a Correlation": a case study of toxicity of (Benzo-)triazoles on algae. *Mol. Inform.*, **31**, 817–835.

227. Gramatica, P., Chirico, N., Papa, E. *et al.* (2013) QSARINS: a new software for the development, analysis, and validation of QSAR MLR models. *J. Comput. Chem.*, **34**, 2121–2132.

228. Xi, L.L., Sun, H.J., Li, J.Z. *et al.* (2010) Prediction of infinite-dilution activity coefficients of organic solutes in ionic liquids using temperature-dependent quantitative structure–property relationship method. *Chem. Eng. J.*, **163**, 195–201.

229. Burnham, K.P., Anderson, D.R., and Huyvaert, K.P. (2011) AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav. Ecol. Sociobiol.*, **65**, 23–35.

230. Lukacs, P.M., Burnham, K.P., and Anderson, D.R. (2010) Model selection bias and Freedman's paradox. *Ann. Inst. Stat. Math.*, **62**, 117–125.

231. Kubinyi, H., Folkers, G., and Martin, Y.C. (1998) 3D QSAR in drug design: ligand-protein interactions and molecular similarity – preface. *Perspect. Drug Discov.*, **9–11**, v–vii.

232. ECA. *Guidance on Information Requirements and Chemical Safety Assessment*, http://echa.europa.eu/web/guest/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment (accessed 22 August 2013).

233. IHCP. *European Union Institute for Health and Consumer Protection QSAR Models Webpage*, https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology (accessed 31 March 2016).

234. OECD. *Organisation for Economic Co-operation and Development (OECD) QSAR Project*, http://www.oecd.org/chemicalsafety/testing/oecdquantitativestructure-activityrelation shipsprojectqsars.htm (accessed 22 August 2013).

235. NAFTA. *(Quantitative) Structure Activity Relationship [(Q)SAR] Guidance Document, Technical Working Group on Pesticides*, https://www.epa.gov/sites/production/files/2016-01/documents/qsar-guidance.pdf (accessed 31 March 2016).

236. EPA. *Toxicity Estimation Software Tool (TEST), U. S. Environmental Protections Agency,* https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test (accessed 31 March 2016).

237. He, Y.Y., Liew, C.Y., Sharma, N. *et al.* (2013) PaDEL-DDPredictor: open-source software for PD-PK-T prediction. *J. Comput. Chem.*, **34**, 604–610.

238. Newman, M.C., McCloskey, J.T., and Tatara, C.P. (1998) Using metal-ligand binding characteristics to predict metal toxicity: quantitative ion character-activity relationships (QICARs). *Environ. Health Perspect. Suppl.*, **106**, 1419–1425.

239. Maguna, F.P., Nunez, M.B., Okulik, N.B., and Castro, E.A. (2009) Methodologies QSAR/QSPR/QSTR: current state and perspectives. *Int. J. Chem. Model.*, **1**, 221–243.

240. Greene, N. and Naven, R. (2009) Early toxicity screening strategies. *Curr. Opin. Drug Discov. Devel.*, **12**, 90–97.

241. Andersen, M.E. and Krewski, D. (2009) Toxicity testing in the 21st century: bringing the vision to life. *Toxicol. Sci.*, **107**, 324–330.

242. Zhou, D.-M., Li, L.-Z., Peijnenburg, W.J.G.M. *et al.* (2011) A QICAR approach for quantifying binding constants for metal–ligand complexes. *Ecotoxicol. Environ. Saf.*, **74**, 1036–1042.

243. Ownby, D.R. and Newman, M.C. (2003) Advances in quantitative ion character-activity relationships (QICARs): using metal-ligand binding characteristics to predict metal toxicity. *QSAR Comb. Sci.*, **22**, 241–246.

244. Zamil, S.S., Ahmad, S., Choi, M.H. *et al.* (2009) Correlating metal ionic characteristics with biosorption capacity of *Staphylococcus saprophyticus* BMSZ711 using QICAR model. *Bioresource Technol.*, **100**, 1895–1902.

245. Mendes, L.F., Bastos, E.L., and Stevani, C.V. (2010) Prediction of metal cation toxicity to the bioluminescent fungus *Gerronema viridilucens*. *Environ. Toxicol. Chem.*, **29**, 2177–2181.

246. Mendes, L.F., Zambotti-Villela, L., Yokoya, N.S. *et al.* (2013) Prediction of mono-, bi-, and trivalent metal cation relative toxicity to the seaweed *Gracilaria domingensis* (Gracilariales, Rhodophyta) in synthetic seawater. *Environ. Toxicol. Chem.*, **32**, 2571–2575.

247. Dearden, J.C., Cronin, M.T.D., and Kaiser, K.L.E. (2009) How not to develop a quantitative structure-activity or structure–property relationship (QSAR/QSPR). *SAR QSAR Environ. Res.*, **20**, 241–266.

248. Le, T., Epa, V.C., Burden, F.R., and Winkler, D.A. (2012) Quantitative structure–property relationship modeling of diverse materials properties. *Chem. Rev.*, **112**, 2889–2919.

249. Todeschini, R. *Useful and Unuseful Summaries of Regression models*, http://www.moleculardescriptors.eu/tutorials/T5_moleculardescriptors_models.pdf (accessed 5 October 2013).

250. Consonni, V., Ballabio, D., and Todeschini, R. (2009) Comments on the definition of the Q(2) parameter for QSAR validation. *J. Chem. Inf. Model.*, **49**, 1669–1678.

251. Consonni, V., Ballabio, D., and Todeschini, R. (2010) Evaluation of model predictive ability by external validation techniques. *J. Chemometr.*, **24**, 194–201.

252. Ballabio, D. and Consonni, V. (2013) Classification tools in chemistry. Part 1: linear models. PLS-DA. *Anal. Methods UK*, **5**, 3790–3798.

253. Chirico, N. and Gramatica, P. (2011) Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *J. Chem. Inf. Model.*, **51**, 2320–2335.

254. Chirico, N. and Gramatica, P. (2012) Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. *J. Chem. Inf. Model.*, **52**, 2044–2058.