

Synthetic popularity reflects chemical reactivity

Bartłomiej Kowalczyk^a, Kyle J. M. Bishop^a, Stoyan K. Smoukov^a
and Bartosz A. Grzybowski^{a*}



Large and diverse databases of chemical reactions contain statistically significant information about the propensities of molecules to undergo specific chemical transformations. It is shown that this information can be quantified to reflect reaction thermodynamics/kinetics and can be used to construct primitive (yet accurate) reactivity indices from the counts of reported reactions involving molecules/molecular positions of interest. These indices correlate with frontier orbital (FO) populations or Hammett σ and ρ parameters for a range of reactions involving aromatic substrates. These findings suggest that large chemical databases are not only a historical repository of chemical knowledge but also tools with which one can make useful chemical predictions. Copyright © 2009 John Wiley & Sons, Ltd.

Supplementary material for this paper is available in the online version of this Journal

Keywords: linear free energy relationships; organic chemistry; structure-activity relationship; substituent effect; synthesis design

INTRODUCTION

From the inception of modern chemistry, chemists have strived to correlate reactivities of organic molecules with their structural features and/or electronic characteristics. These efforts resulted in the development of various reactivity measures/indices ranging from empirical correlations (quantitative structure activity relationships, QSAR, linear free-energy relationships, LFER)^[1,2] to molecular-orbital populations derived from quantum-chemical calculations.^[3,4] Several of these methods have proven spectacularly successful in predicting reactivities of specific classes of molecules and/or specific molecular locations (e.g., Woodward–Hoffmann orbital symmetry rules, Fukui's FMOs).^[5–7] Still, most reactivity predictions are not without ambiguity and the development of accurate and general reactivity measures remains one of the main challenges of physical organic chemistry and chemoinformatics. In this paper, we revisit the issue of chemical reactivity from a different perspective that can be termed “knowledge-based”.^[8,9] The starting point of our approach is an observation that every organic reaction carried out is, in essence, an experiment in chemical reactivity and thermodynamics. Therefore, it is reasonable to speculate that large sets of known chemical reactions should contain in themselves statistically significant information about propensities of the substrate molecules to undergo specific chemical transformations. This assumption should be especially true if the analyzed reaction collections are diverse and not biased by the choice of reaction conditions, reagents used, etc. Such diverse reaction sets can be retrieved from comprehensive computer databases (e.g., Beilstein, Gmelin, The Chemical Thesaurus, Current Chemical Reactions) that contain millions of organic transformations reported in literature from the times of Lavoisier (~1780) to the present.^[10] Based on this reasoning, we then formulate the main hypothesis

underlying our work – namely, that reactivities of organic molecules and/or specific molecular positions scale with the counts of known/reported organic reactions involving these molecules/positions (Fig. 1). In other words, we suggest that molecular reactivity should reflect synthetic popularity. To test this hypothesis we analyze several types of “model” aromatic substitutions for which (i) reaction counts are sufficient to permit meaningful statistical treatment and (ii) other types of reactivity indices (here, frontier-orbital populations, and Hammett constants, HC) are well-established. We then show that indices derived from reaction counts correlate with independent FOP or HC results and that the linear relationships between reaction counts and Hammett parameters reflect the underlying reaction mechanisms. Of course, the purpose of these correlations is not to corroborate the FOP or HC approaches but, instead, use them to validate the popularity-based measures. Once validated, the reaction counts can be used to estimate reactivity trends in sets of molecules other than those discussed here (provided that enough reactions have been reported for these molecules in one of the chemical databases). In sum, the logic of our approach is that statistical analysis of the work of past generations of chemists can yield “primitive” measures of molecular reactivity that are of accuracy comparable to much more sophisticated methods such as FOP or HC.

Figure 1 illustrates the overall logic of our approach. Consider a molecule in which positions *A* and *B* can be loci of a specific type,

* Correspondence to: B. A. Grzybowski, Department of Chemistry and Chemical Engineering, Northwestern University, Evanston, 60208 IL, USA.
E-mail: grzybor@northwestern.edu

^a B. Kowalczyk, K. J. M. Bishop, S. K. Smoukov, B. A. Grzybowski
Department of Chemistry and Chemical Engineering, Northwestern University, Evanston, Illinois, USA

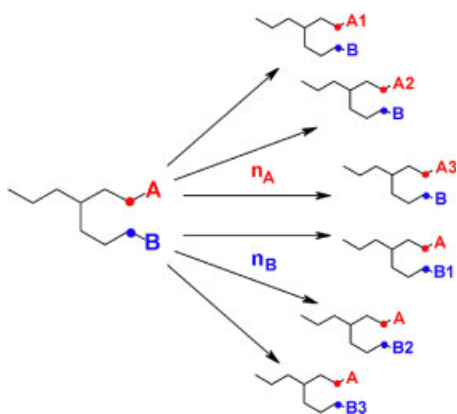


Figure 1. Schematic illustration of estimating reactivities of positions A and B based on the counts of known reactions (n_A and n_B) modifying these positions

T , of chemical reaction (e.g., S_N1 , S_N2 , E1, etc.). To estimate the relative reactivities of A and B, all chemical transformations of type T occurring at/modifying these positions are retrieved from a large, unbiased database of reactions. Here, this database is Beilstein, which is the largest albeit not without omissions-repository of organic reactions reported in the literature from 1779 to the present.^[10,11] Denoting the numbers of retrieved reactions by n_A and n_B , the postulated reactivity indices (or simply, “reactivities”) at positions A and B are then, respectively, $R_A = n_A/(n_A + n_B)$ and $R_B = n_B/(n_A + n_B)$. In the following, we validate this type of an index in several “model” classes of reactions.

ELECTROPHILIC AROMATIC SUBSTITUTIONS

We first consider the “test” case of electrophilic substitutions of monosubstituted benzenes (Fig. 2) with $X = \text{Br}[114,42,733]$, $\text{NH}_2[298,52,906]$, $\text{Cl}[265,125,1291]$, $\text{OH}[1312,70,2669]$, $\text{Me}[920, 519,3166]$, $\text{F}[129,39,654]$, $\text{OMe}[812,138,3391]$, $\text{Et}[107, 67,455]$, $\text{tBu}[53,54,370]$, $\text{Ph}[60,39,663]$, $\text{iPr}[78,43,339]$, $\text{OPh}[50,6, 275]$, $\text{OEt}[51,3,467]$, $\text{CH}_2\text{Cl}[22,11,32]$, where the numbers in the brackets give the reaction counts reported in Beilstein for the *ortho*, n_o^X , *meta*, n_m^X , and *para*, n_p^X , positions. For each substituent, X , reactivities of these positions (R_o , R_m , R_p) toward electrophilic attack are then estimated from the fraction of reaction counts. For

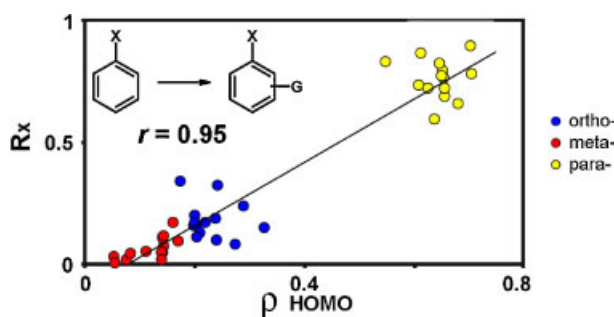


Figure 2. Correlation between the popularity-based reactivity indices and electron populations of HOMO orbitals, ρ_{HOMO} , for electrophilic substitutions of monosubstituted benzenes

example, for $X = \text{Br}$, $n_o^X = 114$, $n_m^X = 42$, and $n_p^X = 733$ such that $R_o = n_o^X/n_{\text{total}}^X = 0.128$, $R_m = n_m^X/n_{\text{total}}^X = 0.047$, and $R_p = n_p^X/n_{\text{total}}^X = 0.825$, where $n_{\text{total}}^X = n_o^X + n_m^X + n_p^X$. While this ordering appears qualitatively correct ($R_p > R_o > R_m$), quantitative validity of the R s can be established only by correlating them with some other, well-established reactivity indices – for example, populations of frontier orbitals (FOs).^[12] This approach is illustrated in Fig. 2 where the values of R_o (blue markers), R_m (red), and R_p (yellow) are plotted against appropriately normalized populations of HOMO orbitals calculated using Gaussian with 6-311+G basis (cf. Supplementary Information). Figure 2 shows that the two measures indeed correlate with the correlation coefficient $r = 0.91$. Interestingly, inclusion of strongly electron withdrawing/deactivating substituents (e.g., NO_2 , CF_3 , or COOH) lowers the value of r . This again, is chemically reasonable given that for such EWG groups, the substitution patterns are not reflected by the FO populations and, instead, it is more appropriate to consider the stability of σ -complex intermediates.^[13] From these results we conclude that the accuracy of R s in predicting reactivity is comparable to that of FO populations.

HAMMETT CONSTANTS AND REACTION MECHANISMS

Since reactivity in electrophilic aromatic substitutions has been studied and understood in great detail, the indices R for this class of reactions have only proof-of-concept value and quite limited practical importance. Popularity-based reactivity measures can be more useful in cases when alternative reactivity descriptors are not so readily available. With this in mind, we consider the Hammett relation whose application typically requires the knowledge of a series of equilibrium or rate constants. If our popularity-based indices correlate with the Hammett constants, relative reactivities within sets of structurally related compounds and even certain aspects of reaction mechanisms could be determined based on the database statistics alone – that is, without the need to perform actual experiments.

In their original formulation,^[14,15] Hammett constants were used to estimate the effects of various *meta* and *para* substituents, X ^[16] on the ionization of benzoic acids, and were defined as $\sigma_X \equiv \log(K_X^B/K_H^B)$, where K_X^B is the ionization constant of an X -substituted benzoic acid in water at 25 °C, and K_H^B is the ionization constant of the unsubstituted benzoic acid. Thus, the Hammett constant provides a measure of the electron-withdrawing or electron-donating character of a given substituent X (i.e., $\sigma_X > 0$ for X more electron-withdrawing than hydrogen; $\sigma_X < 0$ for X more electron-donating than hydrogen).^[17]

When extended to reactions other than the ionization of benzoic acid, the Hammett relation can be written as either $\log(K_X/K_H) = \rho\sigma_X$ or $\log(k_X/k_H) = \rho\sigma_X$, where K_X and K_H are the equilibrium constants (likewise, k_X and k_H are the reaction rate constants) of the given reaction with and without the substituent X , respectively. The proportionality constant, ρ , is specific to a given reaction, and provides a measure of the “sensitivity” of the reaction to substituent effects (Fig. 3). For example, when $\rho < 0$, electron density at the reaction site decreases during the rate-determining step, and the reaction is facilitated by electron-donating groups (EDG, $\sigma_X < 0$) such as NH_2 or OMe . Conversely, when $\rho > 0$, the electron density at the reaction center increases during the rate-determining step; consequently,

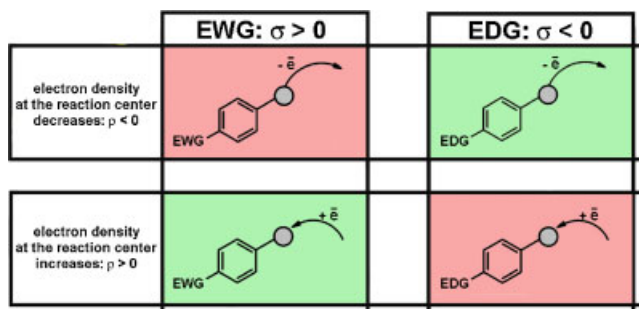


Figure 3. Sigma (σ) and rho (ρ) parameters in Hammett's equation. (top row) Reactions, in which electron density at the reaction center decreases, $\rho < 0$, are facilitated by electron-donating substituents characterized by $\sigma < 0$ (table entry colored green) and are inhibited by electron-withdrawing substituents, $\sigma > 0$ (colored red) (bottom row). If $\rho > 0$ and the electron density at the reaction site increases, reactions are facilitated by electron-withdrawing groups, $\sigma > 0$, and inhibited by electron-donating substituents, $\sigma < 0$

the reaction is facilitated by electron-withdrawing groups (EWGs) characterized by $\sigma_X > 0$ (e.g., CN or NO₂). Importantly, the knowledge of ρ can help to elucidate – or, at least, support – the mechanism of given reaction.

With these preliminaries, we make two related hypotheses. First, that the reactivity indices based on appropriate reaction counts R_X reflect reaction thermodynamics and/or kinetics and are proportional to the relative free energies, $\log(K_X/K_H)$, and/or activation energies, $\log(k_X/k_H)$; therefore, R_X values should correlate with the Hammett constants, σ_X . Second, that if R_X and σ_X indeed correlate, the slope of their dependence is related to ρ , and its sign and magnitude can be used to infer the qualitative features of the reaction mechanism.

Before analyzing specific chemical examples, the R_X parameters need to be properly normalized (cf. Section 1) to take into account different reaction counts for different substrates used to construct each Hammett plot. To begin with, let $n(A_X \rightarrow B_X)$ denote the number of database-reported reactions of type T , in which substrate A , substituted with chemical group X , reacts to form B , also substituted with X . Similarly, $n(A_H \rightarrow B_H)$ stands for “reference” reactions of unsubstituted substrates (i.e., $X = H$). In the spirit of the Hammett relation, R_X should be proportional to the ratio $n(A_X \rightarrow B_X)/n(A_H \rightarrow B_H)$. This expression; however, does not take into account relative database occurrences (“abundances”) of the substrates that might skew the inherent (“true”) reactivity measures. To see this, consider a hypothetical example in which X -substituted substrate A_X is more reactive in some specific reaction T than the unsubstituted A_H . In this case, a correct Hammett-like reactivity measure should clearly be greater than unity, $R_X > 1$. Unfortunately, if A_X is not a popular substrate (say, it is used in only 20 reactions of which 10 are of the type $A_X \rightarrow B_X$) while A_H is used widely (say, in 1000 reactions, of which 100 are of the type $A_H \rightarrow B_H$), the ratio $n(A_X \rightarrow B_X)/n(A_H \rightarrow B_H)$ predicts $R_X = 10/100 = 0.1$. To eliminate this popularity bias, the number of reactions needs to be normalized with respect to the relative database abundances of substrates $R_X = \frac{n(A_X \rightarrow B_X)/n(A_X)}{n(A_H \rightarrow B_H)/n(A_H)}$ where $n(A_X)$ and $n(A_H)$ are the numbers of times A_X and A_H were used as substrates in *any* type of reaction (not only T). For the specific example considered above, the normalized reactivity becomes $R_X = \frac{10/20}{100/1000} = 5$.

Using the R_X indices normalized for database occurrences, we analyzed several classes of common organic reactions. These examples were chosen because for all substituents considered the counts of Beilstein-reported reactions of each type, $n(A_X \rightarrow B_X)$, are relatively large (tens to several hundreds) as are the database occurrences of the substrates ($n(A_X) > 100$ in all cases). In addition, as we will see shortly, these reactions types cover a range of sensitivity constants, ρ . For some reactions, the magnitude of ρ has been reported in the literature thus allowing for direct comparison of our approach with experiment. In all cases, the values of σ_X against which popularity indices were plotted were taken from reference [18].

Formation of diazo-compounds from substituted anilines (Fig. 4)

Formation of diazo-compounds from substituted anilines was analyzed for $X = \text{NH}_2$ [21,2577], OH[26,2703], OMe[92,8931], Me[140,11912], F[30,2089], I[15,966], Br[57,3607], Cl[90,8332], COOH[58,2486], SO₂NH₂[38,1520], CN[28,1152], NO₂[198,5466], and H[364,37247], where the square brackets give the numbers of reactions $n(A_X \rightarrow B_X)$ and $n(A_X)$ for each substituent.^[19] Figure 4(a) gives the plot of R_X versus σ_X where the correlation coefficient is $r = 0.74$ ($r = 0.79$ for *meta*-substituted anilines, Fig. 4(b); $r = 0.84$, for *para*-substituted anilines, Fig. 4(c)). By the arguments developed earlier, the positive slope of this dependence suggests that $\rho > 0$ and the reaction is facilitated by EWGs ($\sigma_X > 0$). This, in fact, agrees with the well-established two-step reaction mechanism^[20] outlined in Fig. 4(d). Here, the

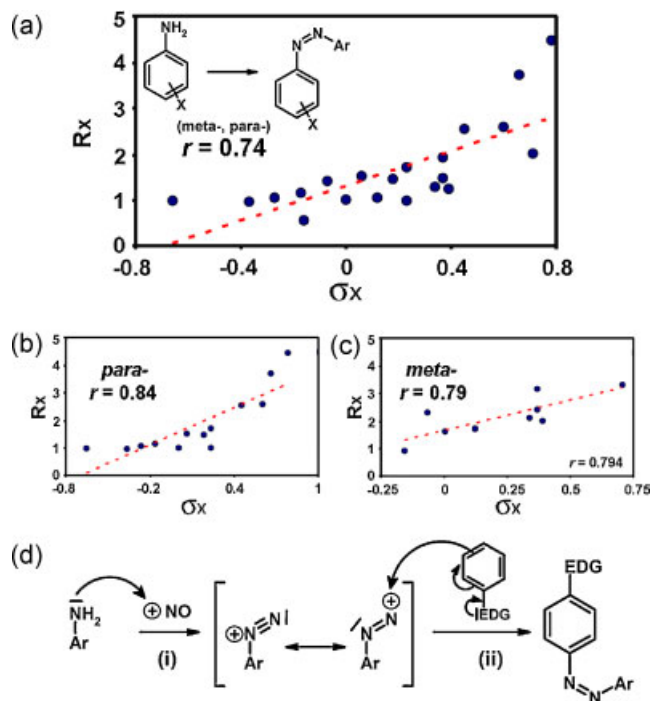


Figure 4. (a) Correlation between Hammett's σ_X and popularity index, R_X , for the formation of diazo compounds from substituted anilines (either *meta* or *para*) (b,c) Correlations for *meta* and *para* substituents. Reaction counts for these graphs are provided in the Supplementary Information. Also, only substituents with counts greater than 5 are included. (d) Reaction mechanism

first step is the formation of diazonium cation by reaction of aniline with nitrosonium cation. Although in highly acidic environments the majority of aniline exists as a salt (e.g., hydrochloride), only the free amine is reactive toward NO^+ electrophile, and the rate-determining step involves the "free" form. Consequently, EWGs that stabilize "free" aniline (with electron density on nitrogen atom higher than in aniline hydrochloride) facilitate the reaction (literature values for this step are $\rho \sim 2.3\text{--}2.6$).^[21,22] In the second step, diazonium salt is attacked by an electron pair of another aromatic molecule (typically, aniline or phenol) which, again, increases the electron density at the reaction center. Similar to the first reaction step, EWGs facilitate this transformation, for which the literature-reported values of ρ are also highly positive.^[23] Interestingly, the magnitude of ρ from the R_X versus σ_X slope is around 2 – that is, close to the literature reported values.

Stille cross-coupling of *para*-substituted bromobenzenes (Fig. 5)

Stille cross-coupling of *para*-substituted bromobenzenes was studied for $X = \text{NH}_2$ [34,2861], OH [34,1165], OMe [225,1931], tBu [18,372], Me [133,1345], Ph [42,461], F [50/557], CO-Me [124,1442], COOH [38,434]; COOMe [57,303], COOEt [48,223], CN [146,899], NO_2 [103,713], and H [404,4785]). Figure 5(a) shows that R_X and σ_X are positively correlated ($r = 0.74$) suggesting that this reaction should be characterized by a positive sensitivity constant, $\rho > 0$. Although the details of the mechanism are not fully understood and may change with different reaction conditions,^[24] transmetalation (Fig. 5(b)) is thought to be the rate determining step with most substrates.^[25,26] Here, the aryl group of organostannate nucleophile reacts at the activated palladium reaction center of bromobenzene, and the electron density on palladium increases upon substitution of the halogen by aryl. Therefore, the reaction should be facilitated by EWGs ($\rho > 0$), in agreement with the popularity analysis. For this reaction, literature sources do not provide exact magnitudes of ρ though they mention it should be positive.

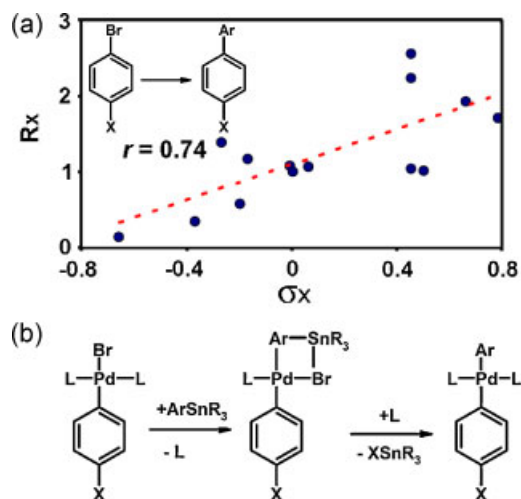


Figure 5. (a) Correlation between R_X and σ_X for the Stille coupling reaction. (b) Mechanism of transmetalation (rate determining) step in the Stille reaction. This figure is available in colour online at www.interscience.wiley.com/journal/poc

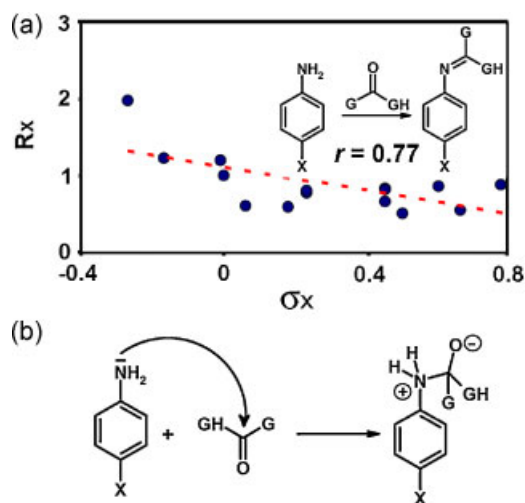


Figure 6. (a) The plot of R_X versus σ_X for the formation of imines via reaction of anilines with aldehydes/ketones. (b) Rate determining step. This figure is available in colour online at www.interscience.wiley.com/journal/poc

Formation of imines from *p*-substituted anilines and carbonyl electrophiles (aldehydes/ketones) (Fig. 6)

Formation of imines from *p*-substituted anilines and carbonyl electrophiles (aldehydes/ketones) was analyzed for substituents $X = \text{OMe}$ [309,8366], Me [231,10147], Ph [13,579], F [21,1860], I [9,824], Cl [96,6695], Br [45,2861], COOH [23,1890], COMe [9,967], COOMe [9,584], COOEt [21,1378], SO_2NH_2 [23,1437], CN [8,789], NO_2 [66,3909] and H [724,38740]. Figure 6(a) shows that the correlation between R_X versus σ_X is negative with the slope of the regression line, -0.77 , corresponding to a negative value of ρ . This trend suggests that the reaction should be facilitated by electron-donating groups (EDGs, $\sigma_X < 0$). Indeed, formation of imines from amines and ketones/aldehydes starts with nucleophilic attack of the amine on the carbonyl group of the ketone/aldehyde. During this rate-determining step,^[27] EDGs increase the nucleophilicity of aniline and stabilize positive charge developed after addition to the carbonyl group of an aldehyde/ketone (Fig. 6(b)). Consequently, the reaction is characterized by $\rho < 0$, as predicted by the popularity analysis. Notably, the slope of the predicted trend, $\rho = -0.77$ is relatively close to the literature reported values for similar types of transformation ($\rho = -1.20$ for reaction of *p*-substituted anilines with formaldehyde).^[21,28] Also for the reverse reaction, i.e., hydrolysis of imines, experimental studies^[21,29] give the slopes of similar magnitudes but opposite sign, $\rho = 0.42\text{--}0.68$, as might be expected.

Other examples

Correlations between R_X and σ_X also hold for reaction types other than those described in the previous sub-sections. Some examples for which database statistics are most abundant include: (i) Heck and/or Suzuki type coupling of *p*-substituted bromobenzenes with alkenes ($r = 0.70$; $\rho = 0.80$ vs. literature values of $\rho = 0.2\text{--}5.0$ depending on the catalysts used^[30]), (ii) hydrolysis of *p*-substituted benzoates ($r = 0.68$; $\rho = 2.7$, compared to the reported^[21,31–33] $\rho = 1.9\text{--}2.8$), and (iii) hydrolysis of *p*-substituted benzonitriles ($r = 0.58$; $\rho = 2.6$ compared to the reported^[34,35] $\rho = 1.8\text{--}2.5$). From these and other examples we

studied, we conclude that the approach based on popularity indices yields predictions of quality comparable to experimental Hammett analyses.

ANALYSES BASED ON SPARSE STATISTICS

The last question we consider concerns the reliability of the trends in situations when there are few database entries from which to calculate the reactivity indices. Consider two cases where the database statistics are sparse. The first example deals with the synthesis of coumarins by aldol condensation of salicylaldehyde and acetic anhydride, for which the substituents and reaction counts are $X = \text{OH}[6,748]$, $\text{OMe}[4,415]$, $\text{Me}[2,64]$, $\text{Br}[1,32]$, $\text{NO}_2[4,37]$, $\text{H}[10,5280]$. Figure 7(a) shows that despite low counts of $n(A_X \rightarrow B_X)$, the popularity indices correlate very well with the Hammett's constants ($r = 0.89$) and yield positive value of ρ (as expected, since in the first and rate determining step the nucleophilic attack of deprotonated anhydride anion on the carbonyl group of aldehyde develops negative charge at the reaction center, which is stabilized by EWGs and destabilized by EDGs). Such strong correlations, however, are by no means typical when dealing with sparse statistics. As an example, consider free-radical halogenation (bromination and chlorination) at the benzylic position of substituted toluenes. For this reaction type, Beilstein yields $X = \text{OMe}[9,519]$, $\text{Me}[13,1549]$, $\text{Cl}[5,591]$, $\text{Br}[5,1345]$, $\text{COOH}[8,612]$, $\text{CN}[5,659]$, and $\text{H}[512,6993]$. The plot of Figure 7(b) has the dependence of R_X on σ_X . Although the slope is negative, $\rho < 0$, as expected (since EWGs stabilize the unpaired electron and make the radical less reactive), the correlation is very poor with $r = 0.35$ (the calculated and experimental^[36] ρ s are -0.62 and -1.37 , respectively). Clearly, one must be cautious when drawing conclusions from trends based on few reaction counts.

As a rule of thumb, the number of reactions of type T should be greater than ~ 10 (for each substituent) to give reliable reactivity estimates. To see this, note that the reactivity measure, R_X , depends on the fractions, f , of total reactions that are of type T for

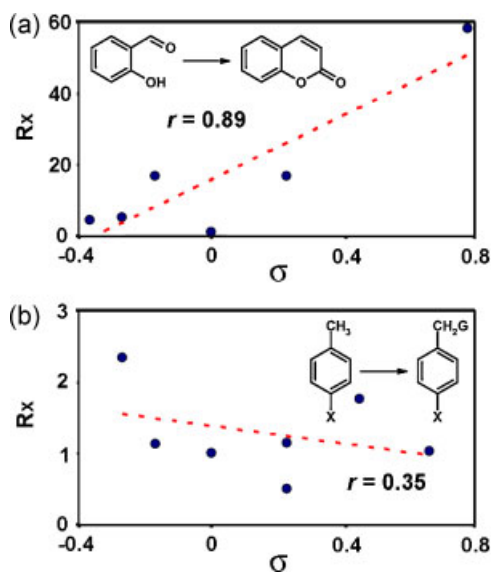


Figure 7. Dependencies of R_X versus σ_X for (a) coumarin synthesis via aldol condensation (b) free-radical halogenation of *para*-substituted toluenes. This figure is available in colour online at www.interscience.wiley.com/journal/poc

both the X -substituted and the unsubstituted reagents. In practice, this fraction is estimated from a finite sample of reactions obtained from the database and containing n_T reactions of type T in a total of n reactions. Based on this sample, we may infer the most probable ("true") value of f , $f_0 = n_T/n$, as well as the standard deviation, s , for our estimate by assuming a binomial likelihood function (i.e., assuming n_T follows a binomial distribution with parameters, n and f).^[37] To obtain a reasonable estimate of f , the standard deviation should be small relative to the mean expectation – that is, $s/\langle f \rangle \ll 1$ or $\sqrt{\frac{(n-n_T+1)}{(n+3)(n_T+1)}} \ll 1$, which is satisfied for $n_T > 10$ regardless of the total reaction count, n . Furthermore, when estimating ρ , it is best to include "reliable" data points from as many substituents, X , as possible in order to accurately assess the correlation (positive, negative, or none) with the Hammett constants.

SUMMARY

The examples we discussed demonstrate that synthetic popularity statistics can be used to estimate molecular reactivity of compounds stored in large chemical databases such as Beilstein. Next time the reader is interested in a reactivity pattern of a known molecule of class of molecules, he or she might want to consider the simple counting of database-reported reactions, which should give reactivity estimates comparable to higher-end methods. Of course, this "knowledge-based" approach is sensitive to the database abundances of reactions/substituents that can be included in the reactivity analysis. When the available statistics are sparse, the reactivity predictions are not reliable. This is most limiting for the molecules that are structurally more complex (and, unfortunately, interesting!) than the aromatics we considered here – for complex molecules, relatively few reactions are reported (for $\text{MW} > \sim 150$, the numbers of reactions in which a molecule is used as a substrate drops rapidly as MW^{-3} , as shown in Reference 10 for connectivity statistics).

Fortunately, recent observation that the "universe" of reported reactions increases exponentially with time^[10,11] tells us that the statistics can only get better as time passes. Databases other than Beilstein could also be used, though they are usually costly (e.g., SPRESI with *ca.* three millions reactions costs approximately \$100 000) or proprietary. Last but not least, it is somewhat uplifting to realize that no reaction we, chemists, publish is useless – even if it does not transform chemistry *en large*, it can contribute to our understanding of chemical reactivity through database-derived popularity measures.

REFERENCES

- [1] J. Devillers, *Comparative QSAR*, Taylor & Francis, New York, **1998**.
- [2] L. B. Kier, L. Hall, *Molecular Connectivity in Structure Activity Relationships*, Research Studies Press, Letchworth, **1986**.
- [3] R. Ponec, L. Amat, R. Carbo-Dorca, *J. Phys. Org. Chem.* **1999**, *12*, 447–454.
- [4] M. Karelson, V. S. Lobanov, A. R. Katritzky, *Chem. Rev.* **1996**, *96*, 1027–1043.
- [5] R. B. Woodward, R. Hoffmann, *J. Am. Chem. Soc.* **1965**, *87*, 395–397.
- [6] R. Hoffman, *Angew. Chem., Int. Ed.*, **2004**, *43*, 6586–L 6590.
- [7] K. Fukui, H. Fujimoto, *Frontier Orbitals and Reaction Paths: Selected Papers of Kenichi Fukui*, World Scientific Publishing Co., Singapore, **1998**.

- [8] B. A. Grzybowski, A. V. Ishchenko, C. Y. Kim, G. Topalov, R. Chapman, D. W. Christianson, G. M. Whitesides, E. I. Shakhnovich, *Proc. Natl. Acad. Sci.* **2002**, *99*, 1270–1273.
- [9] B. A. Grzybowski, A. V. Ishchenko, J. Shimada, E. I. Shakhnovich, *Acc. Chem. Res.* **2002**, *35*, 261–269.
- [10] M. Fialkowski, K. J. M. Bishop, V. A. Chubukov, C. J. Campbell, B. A. Grzybowski, *Angew. Chem., Int. Ed.* **2005**, *44*, 7263–7269.
- [11] K. J. M. Bishop, R. Klajn, B. A. Grzybowski, *Angew. Chem., Int. Ed.* **2006**, *45*, 5348–5354.
- [12] I. Fleming, *Frontier Orbitals and Organic Chemical Reactions*, VCH, New York, **1976**.
- [13] F. A. Carey, R. J. Sundberg, *Advanced Organic Chemistry*, Plenum Press, New York, **1993**.
- [14] L. P. Hammett, *Chem. Rev.* **1935**, *17*, 125–L 136.
- [15] L. P. Hammett, *J. Am. Chem. Soc.* **1937**, *59*, 96–103.
- [16] *Ortho*- substituents do not follow Hammett's equation because of additional steric effects.
- [17] The net electronic effect of substituent *X* may be considered as the sum of two distinct effects: induction and resonance. The effect of resonance is greater for *para* than for *meta* substituents due to better conjugation of electron pairs in the *para* position. In contrast, inductive effects, which result from the polarization of bonds, are stronger for *meta* than for *para* substituents, which are more distant from the reaction site. Consequently, the Hammett constant σ depends not only on the particular substituent but also on the substituent's position and is measured separately for *meta* and *para* positions (denoted as σ_m and σ_p respectively).
- [18] C. Hansch, A. Leo, R. W. Taft, *Chem. Rev.* **1991**, *91*, 165–195.
- [19] For other *X*'s the reaction counts are below 5.
- [20] A. F. Hegarty, The chemistry of diazonium and diazo groups, In *The Chemistry of Functional Groups*, (Ed.: S. Patai) Wiley, New York, **1978**.
- [21] H. H. Jaffe, *Chem. Rev.* **1953**, *53*, 191–261.
- [22] S. Ueno, T. Suzuki, *J. Soc. Chem. Ind. (Japan)* **1933**, *36B*, 615.
- [23] H. Zollinger, *Azo and Diazo Chemistry – Aliphatic and Aromatic Compounds*, Interscience Publishers, London, **1961**.
- [24] A. L. Casado, P. Espinet, *J. Am. Chem. Soc.* **1998**, *120*, 8978–8985.
- [25] P. Espinet, A. M. Echavarren, *Angew. Chem., Int. Ed.* **2004**, *43*, 4704–4734.
- [26] V. Farina, B. Krishnan, *J. Am. Chem. Soc.* **1991**, *113*, 9585–9595.
- [27] M. B. Smith, J. March, *March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*, (6th edn), John Wiley, New York, **2007**.
- [28] O. C. M. Davis, F. W. Rixon, *J. Chem. Soc.* **1915**, *107*, 728.
- [29] A. V. Willii, R. E. Robertson, *Can. J. Chem.* **1953**, *31*, 361.
- [30] S. Proch, Y. Mei, J. M. R. Villanueva, Y. Lu, A. Karpov, M. Ballauff, R. Kempe, *Adv. Synth. Catal.* **2008**, *350*, 493.
- [31] K. Ingold, S. Natshaw, *J. Chem. Soc.* **1936**, 222.
- [32] E. Tommila, *Ann. Acad. Sci. Fennicae* **1942**, A69, No. 3.
- [33] D. P. Evans, J. J. Gordon, H. B. Watson, *J. Chem. Soc.* **1937**, 1430.
- [34] K. Bowden, S. P. Hiscocks, M. K. Reddy, *J. Chem. Soc., Perkin. Trans. 2* **1997**, 1133.
- [35] R. Prihod'ko, M. Sychev, I. Kolomitsyn, P. J. Stobbelaar, E. J. M. Hensen, R. A. van Santen, *Microporous Mesoporous Mater.* **2002**, *56*, 241.
- [36] C. Kooymaen, R. Vanh Helden, A. F. Bickel, *Koninkl. Ned. Akad. Wetenschap. Proc.* **1953**, *56B*, 75.
- [37] D. S. Sivia, J. Skilling, *Data Analysis: A Bayesian Tutorial*, Oxford University Press, New York, **2006**.