Use of Calculated Quantum Chemical Properties as Surrogates for Solvatochromic Parameters in Structure–Activity Relationships

CHRISTOPHER J. CRAMER*

Department of Chemistry and Supercomputer Institute, University of Minnesota, 207 Pleasant Street SE, Minneapolis, Minnesota 55455

GEORGE R. FAMINI*

U.S. Army Edgewood Research, Development, and Engineering Center, SCBRD-RT (Building E-3160), Aberdeen Proving Ground, Maryland 21010

ALFRED H. LOWREY

The Laboratory for Structure and Matter, U.S. Naval Research Laboratory, Washington, D.C. 20375

Received May 28, 1993

Introduction

The exponential growth of computer power in the last three decades has had a profound impact on the physical sciences. Chemistry is no exception in this regard: virtually every facet of our wide-ranging field has taken advantage of constantly expanding computational power.[1] In particular, the calculation of atomic and molecular properties from first principles by approximate solution of the fundamental equations of quantum mechanics is now commonplace.[2,3] This practice goes by a number of names at present, e.g., computational chemistry, molecular modeling, etc. For ease of discussion, we will refer to such de novo prediction of molecular properties simply as "theory".

Moreover, we choose to identify three distinct, albeit overlapping, philosophies with which theory may be employed. (1) Theory may be applied post facto to a situation where some ambiguity exists in the intepretation of experimental results. For example, photolysis in an inert matrix may lead to a single product species as analyzed by spectroscopy. However, the identity of this unique product may not be obvious given a number of plausible possibilities. Ab initio calculation of the energies and spectra for *all* postulated products provides an opportunity for comparison.[4] (2) Theory may be employed in a simultaneous fashion to improve the design and progress of an experimental program. Continuing the above analogy, a priori calculation of spectra for plausible products may assist in choosing experimental parameters to permit the observation of minor components which might otherwise be missed in complicated mixtures.[4] (3) Finally, theory may be used to predict properties which might be especially difficult, dangerous, or even impossible to measure experimentally. In the difficult category are such data as rate constants for the reactions of trace, upper-atmospheric constituents which play an important role in the ozone cycle.[5] As for dangerous aspects, theoretical prescreening of a series of toxic or explosive compounds can eliminate the need for a complete (and probably more costly) experimental screen.[6] Finally, while there is some ontological question about the nature of properties which are not measurable, it is in general no more difficult, and often quite useful, to apply theory to transition states or even nonstationary points on molecular hypersurfaces.[7] Ultimately, of course, all

Christopher J. Cramer was born in Freeport, IL, on September 23, 1961. He holds an A.B. summa cum laude in mathematics and chemistry from Washington University in St. Louis and a Ph.D. in chemistry from the University of Illinois. After postdoctoral research with the United States Army, he joined the faculty of the University of Minnesota in 1992. His research interests include the theoretical modeling of chemical and biological properties and reactivity.

George R. Famini is a senior research chemist at the U.S. Army Edgewood Research, Development, and Engineering Center, where he has specialized in the applications of computational chemistry and molecular modeling techniques to problems in chemical defense. He graduated from Towson State University (B.S.), University of Delaware (M.S.), and Lehigh University (Ph.D.).

Alfred H. Lowrey is a research chemist at the Laboratory for the Structure of Matter. His primary research interest is prediction of small-molecule structures using the techniques of computational chemistry and gas-phase electron diffraction. He holds a B.S. in chemistry from Stanford and a Ph.D. in theoretical chemistry from Yale.

(1) *Supercomputer Research in Chemistry and Chemical Engineering*; Jensen, K. F., Truhlar, D. G., Eds.; American Chemical Society: Washington, DC, 1987.

(2) (a) *Chemical Applications of Atomic and Molecular Electrostatic Potentials*; Politzer, P., Truhlar, D. G., Eds.; Plenum: New York, 1981. (b) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986. (c) Dykstra, C. E. *Ab Initio Calculation of the Structures and Properties of Molecules*; Elsevier: Amsterdam, 1988.

(3) (a) Dykstra, C. E.; Augspurger, J. D.; Kirtman, B.; Malik, D. J. *Rev. Comput. Chem.* **1990**, *1*, 83. (b) Politzer, P.; Murray, J. S. *Rev. Comput. Chem.* **1991**, *2*, 273.

(4) A limited set of examples includes the following. (a) ESR: Cramer, C. J. *J. Mol. Struct. (THEOCHEM)* **1991**, *235*, 243. (b) IR: Hess, B. A., Jr.; Schaad, L. J.; Cársky, P.; Zahradník, R. *Chem. Rev.* **1986**, *86*, 709. (c) NMR: Wiberg, K. B.; Hadad, C. M.; Sieber, S.; Schleyer, P. v. R. *J. Am. Chem. Soc.* **1992**, *114*, 5820. (d) UV: Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *J. Phys. Chem.* **1992**, *96*, 135.

(5) (a) JPL/NASA. *Chemical Kinetics and Photochemical Data for Use in Stratospheric Modeling, NASA Kinetics Panel Evaluation 7*; Publication 85-37; JPL: Pasadena, 1985. (b) Atkinson, R. *Chem. Rev.* **1986**, *86*, 69.

(6) For examples of the economic benefits of theoretical screening, see: Boyd, D. B. *Rev. Comput. Chem.* **1990**, *1*, 355.

0001-4842/93/0126-0599$04.00/0

such predictions must be corroborated by some sort of relevant experiment: the primary intent of the last philosophy is to permit the design of that experiment to be as safe, efficient, and cost-effective as possible. In this regard, our own interest has been primarily in the area of using theory to predict the properties of compounds with potentially unique biological activity.

Structure–Activity Relationships. A variety of approaches are now being used to predict and understand biological, pharmacological, physical, and chemical properties. Foremost among these have been different implementations of quantitative structure–activity relationships (QSAR).[8] The basic hypothesis of QSAR is that microscopic structural features of a chemical compound can be related to its macroscopic properties. QSAR has been applied in a number of ways, ranging from pattern recognition[9] and principal component analysis[10] to the extrathermodynamic methods of Hansch.[11] In general, a QSAR analysis involves a regression equation that correlates *selected* microscopic features of a set of chemical compounds with a specific property. This approach has been extremely successful in finding relationships between a number of biological and toxicological properties and various aspects of molecular structure. Indeed, an ever-increasing number of descriptors (structural features) are being used in QSAR correlations. This proliferation, as well as the fact that most QSAR regressions have been class-dependent, has made comparison between different regression equations difficult.[8c]

Linear Free Energy Relationships. QSAR is predated by another technique employing the identical mathematical formalism, albeit with only a single descriptor. Thus, connections between structure and reactivity were reviewed early on by Burkhardt[12] and Hammett.[13] Thereafter, exploring the effect of aromatic substitution on the acidity of a number of benzoic acids, Hammett proposed the equation that now bears his name,[14]

$$\log(K_i/K_0)_j = \sigma_i\rho_j \qquad (1)$$

where K_0 is the equilibrium constant for the aqueous ionization of unsubstituted benzoic acid, K_i is the observed acidity for substitution pattern i, and σ_i is derived by assigning ρ to be 1.00 for this reference set of equilibria. For other similar reactions j, regressing observed equilibrium or rate constants for a few substituted cases on their appropriate σ_i values allows for the derivation of ρ_j for that process, and thus the prediction of equilibrium or rate constants for *all* substitution patterns having a known σ_i. This and many similar so-called linear free energy relationships (LFER) have been extensively reviewed.[15,16]

Linear Solvation Energy Relationships. A cross between QSAR and LFER, the linear solvation energy relationship (LSER) was developed by Kamlet, Taft, and co-workers to explain solvent effects on various free energy based properties.[17] Thus, solvents were characterized by constants which described their ability to accept (β) and donate (α) hydrogen bonds, and their dipolarity/polarizability (π^*). The individual values were originally obtained by measuring UV spectral shifts for judiciously selected dyes in the various solvents, and thus they have come to be known as solvatochromic descriptors. These parameters then found use in the equation

$$\log(\gamma) = c_0 + c_1\alpha + c_2\beta + c_3\pi^* \qquad (2)$$

where γ is some solvent-dependent (often free energy based) property, and the coefficients c_i are determined from a multiple linear regression of measured values against the solvatochromic parameters. Again, the objectives are both to understand how solvent affects γ and to permit the *prediction* of γ in *any* solvent for which the solvatochromic parameters are available. This approach and common extensions which consider such other factors as solvent cohesive pressure[18] and coordinate covalency[17b,19] have also been exhaustively reviewed.[20]

Equation 2 was originally designed to correlate the behavior of an individual solute in multiple solvents; however, Kamlet, Taft, Abraham, and others went on to apply this formalism in a reverse sense, i.e.,

$$\log(\gamma) = c_0 + c_1\alpha_2 + c_2\beta_2 + c_3\pi_2^* \qquad (3)$$

where the subscript 2 on the various solvatochromic descriptors now indicates that they are specific to the *solute*.[17c,21] Thus, eq 3 is used to correlate some property exhibited by a class of solutes in a single solvent against the solvatochromic descriptors specific to each solute. In general, a molar or intrinsic volume term is also added, and this approach has found particular use in the calculation of solubilities and partition coefficients.[22] We have now come full circle, in the sense that this approach is identical to QSAR, except that the regression variables are *chemically* derived as opposed to *structurally* derived.

(7) See, for instance, the following. (a) For modeling barrier heights: Dorigo, A. E.; Houk, K. N. *J. Am. Chem. Soc.* **1987**, *109*, 3698. (b) For modeling isotope effects: Liu, Y.-P.; Lynch, G. C.; Truong, T. N.; Lu, D.; Truhlar, D. G.; Garrett, B. C. *J. Am. Chem. Soc.* **1993**, *115*, 2408. (c) For calculating vibrationally averaged expectation values: Cramer, C. J. *J. Org. Chem.* **1991**, *56*, 5229.

(8) For comprehensive reviews, see: (a) Weinstein, H.; Osman, R.; Green, J. P.; Topiol, S. Reference 2a, p 309. (b) Hall, L. H.; Kier, L. B. *Rev. Comput. Chem.* **1991**, *2*, 367. (c) Bersuker, I. B.; Dimoglo, A. S. *Rev. Comput. Chem.* **1991**, *2*, 423. (d) Hansch, C. *Acc. Chem. Res.* **1993**, *26*, 147.

(9) (a) Free, S. M.; Wilson, J. W. *J. Med. Chem.* **1964**, *7*, 395. (b) Fujita, T.; Ban, T. J. *J. Med. Chem.* **1971**, *14*, 148.

(10) (a) Chastrette, M.; Rajzmann, M.; Chanon, M.; Purcell, K. F. *J. Am. Chem. Soc.* **1985**, *107*, 1. (b) Martin, Y. C.; Holland, J. B.; Jarboe, C. H.; Plotnikoff, N. J. *J. Med. Chem.* **1974**, *17*, 409.

(11) Hansch, C.; Fujita, T. J. *J. Am. Chem. Soc.* **1964**, *86*, 1616.

(12) Burkhardt, G. N. *Nature (London)* **1935**, *17*, 684.

(13) Hammett, L. P. *Chem. Rev.* **1935**, *17*, 125.

(14) Hammett, L. P. *J. Am. Chem. Soc.* **1937**, *59*, 96.

(15) Exner, O. In *Correlation Analysis of Chemical Data*; Shorter, J., Ed.; Plenum: New York, 1988.

(16) (a) Jaffé, H. H. *Chem. Rev.* **1953**, *53*, 191. (b) Exner, O. *Prog. Phys. Org. Chem.* **1990**, *18*, 129.

(17) (a) Abboud, J.-L. M.; Kamlet, M. J.; Taft, R. W. *J. Am. Chem. Soc.* **1977**, *99*, 8325. (b) Kamlet, M. J.; Taft, R. W. *Prog. Org. Chem.* **1983**, *48*, 2877. (c) Kamlet, M. J.; Abraham, M. H.; Doherty, R. M.; Taft, R. W. *J. Am. Chem. Soc.* **1984**, *106*, 464.

(18) Abraham, M. H.; Doherty, R. M.; Kamlet, M. J.; Harris, J. M.; Taft, R. W. *J. Chem. Soc., Perkin Trans. 2* **1987**, *913*, 1097.

(19) (a) Brady, J. E.; Carr, P. W. *Anal. Chem.* **1982**, *54*, 1751. (b) Brady, J. E.; Carr, P. W. *J. Phys. Chem.* **1982**, *86*, 3053.

(20) (a) For reviews on this and other LSER formulations, see: Reichardt, C. *Solvents and Solvent Effects in Organic Chemistry*, 2nd ed.; VCH: New York, 1990; Chapter 7. (b) Gajewski, J. J. *J. Org. Chem.* **1992**, *57*, 5500.

(21) Kamlet, M. J.; Taft, R. W. *Acta Chem. Scand.* **1986**, *B40*, 619.

(22) (a) Taft, R. W.; Abraham, M. H.; Famini, G. R.; Doherty, R. M.; Abboud, J.-L. M.; Kamlet, M. J. *J. Pharm. Sci.* **1985**, *74*, 807. (b) Leahy, D. E.; Carr, P. W.; Pearlman, R. S.; Taft, R. W.; Kamlet, M. J. *Chromatographia* **1986**, *21*, 473. (c) Kamlet, M. J.; Taft, R. W.; Famini, G. R.; Doherty, R. M. *Acta Chem. Scand.* **1987**, *B41*, 589. (d) Abraham, M. H.; Whiting, G. S.; Fuchs, R.; Chambers, E. J. *J. Chem. Soc., Perkin Trans. 2* **1990**, 291.

Theoretical Linear Solvation Energy Relationships. The LSER methodology has been very successful in correlating solvatochromic descriptors with a wide range of properties. Like QSAR, its chief virtue is that the LSER coefficients provide *chemical* insight into the specific molecular properties or interactions which impact most strongly on the process under study. However, the methodology must be applied a posteriori to the experimental measurement of the required solvatochromic descriptors. Clearly, the regression can only be used in a predictive mode for the set of all molecules for which these descriptors exist. It is at this point that theory may make a substantial contribution.

One may envision two possibilities in this regard. In order to tie in directly with eq 3, one might attempt to correlate computationally derived structural and/or electronic properties with the solute-specific solvatochromic parameters, and thereby provide a prescription for predicting them from first principles. Such an approach has been explored by Lewis.[23] The alternative is to maintain the *formalism* of LSER, but to abandon its empirical descriptors and replace them with others calculable by theory, i.e., a theoretical linear solvation energy relationship (TLSER).

The use of theoretically or computationally derived parameters in QSAR and QSAR-like correlations has seen much exploration.[8,24] These descriptors have commonly been used in conjunction with empirically derived descriptors, e.g., the octanol–water partition coefficient.[25] Ford and Livingstone have recently summarized some of the inherent advantages of computationally derived descriptors by comparison to those measured experimentally.[26]

We have chosen to apply the philosophy of LSER with the greater flexibility and applicability afforded by the exclusive use of quantum chemically derived descriptors. Thus, we have developed a new, theoretical set of parameters capable of correlating a wide variety of properties.[27] While these parameters are derived solely from computational methods, we have deliberately chosen them to have an obvious correspondence to the LSER parameters. The most general TLSER is thus

$$\log(\gamma) = c_0 + c_1 V_{mc} + c_2 \pi^* + c_3 \epsilon_\alpha + c_4 \epsilon_\beta + c_5 q^+ + c_6 q^- \quad (4)$$

The TLSER volume term, V_{mc}, is the molecular van der Waals volume calculated according to the method of Hopfinger.[28] The polarizability term, π^*, is derived from the polarization volume computed by the method of Kurtz.[29] Just as with LSER, hydrogen bonding is separated into donor and acceptor components.[30] Since *all* intermolecular interactions can be considered to have varying degrees of covalent and electrostatic components, separate descriptors have been chosen to address each of these two limiting formalisms. The covalent contribution to Lewis basicity, ϵ_β, is taken as the difference in energy between the lowest unoccupied molecular orbital (E_{LUMO}) of water and the highest occupied molecular orbital (E_{HOMO}) of the solute. Thus, smaller values of ϵ_β denote a greater covalent basicity. The electrostatic basicity contribution is denoted by q^-, the magnitude of the most negative atomic partial charge in the molecule. Analogously, the hydrogen bonding acidity is divided into two components: ϵ_α is the energy difference between E_{HOMO} of water and E_{LUMO} of the solute, and q^+ is the magnitude of the most positively charged hydrogen atom in the molecule. Calculation of these descriptors is performed at the neglect of diatomic differential overlap (NDDO)[31] semiempirical level of molecular orbital theory; this level of theory allows the quantum mechanical properties to be calculated reasonably accurately at minimal cost. To date we have employed exclusively the MNDO[32] and AM1[33] Hamiltonians. So long as all calculations are done consistently, there is every likelihood that other semiempirical Hamiltonians or higher ab initio levels of theory would be equally useful.

Necessary Philosophical Ruminations

Multiple linear regression on a given set of descriptors, as exemplified by equations 2–4, may legitimately be regarded as a remarkably old concept: the ancients ascribed all physicochemical properties to the proportions of earth, air, water, and fire which were present in a given substance. Moreover, these four "elements" were regarded as being a *complete* and *orthogonal basis* in a mathematical sense, i.e., they were each unique and by appropriate linear combination spanned the space of all observable properties.

It is perhaps not surprising, given this historical precedent, that there has been considerable effort even up to the present to discover measurable descriptors which can be used in similarly minimalist regressions to explain wide varieties of chemical properties and reactivities. Thus, a large number of single descriptors

(23) Lewis, D. F. V. *J. Comput. Chem.* **1987**, *8*, 1084.

(24) (a) Rose, S. L.; Jurs, P. C. *J. Med. Chem.* **1982**, *25*, 769. (b) Loew, G. H.; Poulsen, M.; Kirkjian, E.; Ferrell, J.; Sudhindra, B. S.; Rebagliati, M. *Environ. Health Perspect.* **1985**, *61*, 69. (c) Johnson, J.; Kenley, R. A.; Rynard, C. *Quant. Struct.-Act. Relat.* **1985**, *4*, 172. (d) Lewis, D. F. V. *Prog. Drug Metab.* **1990**, *12*, 205. (e) Hansch, C. In *QSAR: Rational Approaches to the Design of Bioactive Compounds*; Silipo, C., Vittoria, A., Eds.; Elsevier: Amsterdam, 1991. (f) Lewis, D. F. V. *Rev. Comput. Chem.* **1992**, *3*, 173. In addition, for structural descriptors derived from connectivity and graph theoretical techniques, see: (g) Randic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609. (h) Amidon, G. L.; Anik, S. T. *J. Pharm. Sci.* **1976**, *6*, 801. (i) King, R. B. *Chemical Applications of Topology and Graph Theory*; Elsevier: New York, 1983; Vol. 28. (j) Jurs, P. C.; Stouch, T. R.; Czerwinski, M.; Narvaez, J. N. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 296. (k) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Research Studies Press: Letchworth, 1986.

(25) (a) Balducci, R.; Roda, A.; Pearlman, R. S. *J. Solution Chem.* **1989**, *18*, 355. (b) DeVito, S. C.; Pearlman, R. S. *Med. Chem. Res.* **1991**, *1*, 461. (c) Grogan, J.; DeVito, S. C.; Pearlman, R. S.; Korsekwa, K. R. *Chem. Res. Toxicol.* **1992**, *5*, 548. (d) Debnath, A. K.; de Compadre, R. L.; Shusterman, A. J.; Hansch, C. *Environ. Mol. Mutagen.* **1992**, *19*, 53.

(26) Ford, M. G.; Livingstone, D. J. *Quant. Struct.-Act. Relat.* **1990**, *9*, 107.

(27) (a) Famini, G. R. *Using Theoretical Descriptors in Structure Activity Relationships V*; CRDEC-TR-085, U.S. Army Chemical Research, Development, and Engineering Center, Aberdeen Proving Ground, MD, 1989. (b) Wilson, L. Y.; Famini, G. R. *J. Med. Chem.* **1991**, *34*, 1668. (c) Famini, G. R.; Penski, C. A.; Wilson, L. Y. *J. Phys. Org. Chem.* **1992**, *5*, 395. (d) Famini, G. R.; Ashman, W. P.; Mickiewicz, A. P.; Wilson, L. Y. *Quant. Struct.-Act. Relat.* **1992**, *11*, 162. (e) Famini, G. R.; Kassel, R. J.; King, J. W.; Wilson, L. Y. *Quant. Struct.-Act. Relat.* **1991**, *10*, 344.

(28) Hopfinger, A. J. *J. Am. Chem. Soc.* **1980**, *102*, 7126.

(29) Kurtz, H. A.; Stewart, J. J. P.; Dieter, K. M. *J. Comput. Chem.* **1990**, *11*, 82.

(30) Famini, G. R. *Using Theoretical Descriptors in Structure Activity Relationships IV: Molecular Orbital Basicity and Electrostatic Basicity*; CRDEC-TR-88031, U.S. Army Chemical Research, Development, and Engineering Center, Aberdeen Proving Ground, MD, 1988.

(31) Pople, J. A.; Beveridge, D. L. *Approximate Molecular Orbital Theory*; McGraw-Hill: New York, 1970.

(32) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.

(33) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

characterizing a class of molecules/media, which are each derived by the measurement of some single reference process, have been proposed for such properties as solvent polarity,[20a,34] Lewis acidity and basicity,[35] nucleophilicity,[36] solvophobicity,[22c,37] and electronegativity.[38] Of course, each of the individual solvatochromic parameters in eq 2 is an example of this phenomenon. In essence, choices of the appropriate number of descriptors, and the reference processes from which they are derived, have been based primarily on the chemical intuition of individual investigators for a given problem of interest.

Another approach, which is similar in philosophy but quite different in implementation, is to employ principal component analysis (PCA)[39] in a chemometric fashion. PCA identifies the minimum dimensionality (number of "factors") required to correlate one measured property or reactivity (e.g., solvolysis rate in ethanol) with some number of substrate-specific characteristics (e.g., the solute–solvatochromic descriptors, vapor pressure, dipole moment, etc.) Each of the mutually orthogonal factors is formed from a linear combination of the specific characteristics. While this is the mathematical ultimate in reductionism, and thus seems to be in concert with the goals of the empiricists, there has been substantial debate over its relative merit.[21,22a,40] The PCA factors, by virtue of being linear combinants of various characteristics, often are able to correlate a greater "property space"; by that same token, however, they may be quite difficult to interpret in an inductive fashion. The empirical descriptors, on the other hand, clearly lose utility as one applies them to processes quite different from those from which they were derived; however, they have clear physical meaning.

We tend to find this last property of the empirical descriptors compelling, and thus it is no accident that the TLSER descriptors address similar effects as do those for the LSER. However, we should note that we do not view a Procrustean application of a priori reductionism as having any inherent virtue. Certainly it is helpful in understanding chemical processes if the smallest number of important influences can be identified. However, much like the chemometricians, we prefer to winnow them mathematically as opposed to intuitively. Thus, the "default" TLSER represented by eq 4 contains more descriptors than are typical for LSER, and we emphasize that we have no aversion to adding more if necessary and if the new term(s) have a chemically reasonable interpretation. Since the number of theoretical descriptors which can be retrieved from a single calculation is quite large, we are in more of a position to espouse this philosophy than are the empiricists, who must rely on tedious experimental measurements. We ultimately rely upon the regression analysis to eliminate those descriptors which are unimportant. That said, however, it is worth repeating that the objective is for the descriptors to provide insight into the chemistry under study, especially insofar as the sign and magnitude of their regression coefficients are concerned.

The ensuing presentation is designed to both describe the methodology and highlight these philosophical points.

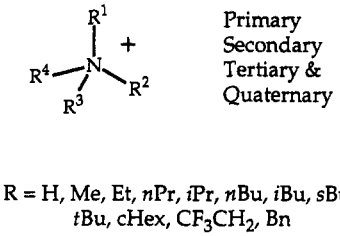


Figure 1. Alkyl ammonium ions exhibiting varying degrees of inhibition of acetylcholinesterase.

Applications

Quantitative Analyses. Equation 4 has been applied to a number of different biological, physicochemical, and spectral properties. In those cases where both TLSER and LSER correlations are available, the TLSER correlation is essentially equal to the LSER correlation. In cases where classical QSAR descriptors have been used (primarily for biological properties), a TLSER analysis has always afforded improved correlation.

One example is provided by the alkyl ammonium ions listed in Figure 1. These ions are observed to inhibit acetylcholinesterase and have been analyzed with classical QSAR techniques.[41] Using eq 4 with MNDO-derived descriptors, we obtain a regression of

$$\log \frac{1}{K_i} = -2.583 - \frac{0.636}{100} V_{mc} + \frac{4.961}{0.1}\pi^* - 2.234q^+ \quad (5)$$

$$N = 44 \qquad R = 0.958 \qquad s = 0.219 \qquad F = 130$$

where N is the number of observations, R is the multiple correlation coefficient, s is the standard error, F measures significance, and the various powders of 10 which appear as denominators in some of the coefficients are used for historical reasons to put all of the theoretical descriptors on a common scale. However, it is not the magnitude of the coefficient but rather its corresponding t-score which indicates its importance in the regression. From examining the t-scores for eq 5, which has an excellent correlation, we find the acetylcholinesterase inhibitory activity to be quite sensitive to polarizability and somewhat less sensitive to both electrostatic acidity and molecular volume. More importantly, use of this equation potentially provides chemical insight into the nature of the interaction between the cations and the enzyme. In particular, recent crystal structures of quaternary ammonium binding proteins indicate that the mode of binding appears to be via cation–π interactions with aromatic residues.[42] This is entirely consistent with the sign and

(34) Review: Buncel, E.; Rajagopal, S. *Acc. Chem. Res.* **1990**, *23*, 226.
(35) (a) Drago, R. S. *J. Chem. Educ.* **1974**, *41*, 300. (b) Persson, I. *Pure Appl. Chem.* **1986**, *58*, 1153.
(36) Winstein, S.; Grunwald, E.; Jones, H. W. *J. Am. Chem. Soc.* **1951**, *73*, 2700.
(37) (a) Leo, A. *J. Chem. Soc., Perkin Trans. 2* **1983**, 825. (b) Menger, F. M.; Venkataram, U. V. *J. Am. Chem. Soc.* **1986**, *108*, 2980.
(38) Luo, Y.-R.; Benson, S. W. *Acc. Chem. Res.* **1992**, *25*, 375 and following comment by Drago, R. S., p 382.
(39) For a review of this and other multivariate analysis methods, see: Jurs, P. C. *Rev. Comput. Chem.* **1990**, *1*, 169.
(40) (a) Sjöström, M., Wold, S. *Acta Chem. Scand., Part B* **1981**, *B35*, 537; **1986**, *B40*, 270. (b) Kamlet, M. J.; Taft, R. W. *Acta Chem. Scand., Part B* **1985**, *B39*, 611.
(41) Takayama, C.; Akamatsu, M.; Fujita, T. *Quant. Struct.-Act. Relat.* **1989**, *8*, 90.
(42) (a) Sussman, J. L.; Harel, F.; Frolow, F.; Oefner, A.; Goldman, A.; Toker, L.; Silman, I. *Science* **1991**, *253*, 872. (b) Satow, Y.; Cohen, G. H.; Padlan, E. A.; Davies, D. R. *J. Mol. Biol.* **1986**, *190*, 593.

Table I. Physicochemical and Biological Properties Which Have Been Correlated with TLSER and TLSER-like Regressions[a]

properties correlated	ref
microtox test	27b
golden orfe toxicity	27b
tadpole narcosis	27b
industrial pollutants	27b
frog muscle inhibition	27b
adsorption on charcoal	27c
Hafkensheid retention indices	27c
octanol/water partition coefficient	27c
organophosphorus hydrolysis rates	27c
pK_a	27c
fentanyl activity at opiate receptor	27d
local anesthetics	27e
dopamine D_1 receptor affinity	46
solubility in supercritical CO_2	47, 48
gas-phase acidity	49
UV–vis spectra of ylides	50
UV–vis spectra of merocyanine dyes	50
solvation enthalpy of *tert*-butyl halides	50
water solubility	50
^{13}C NMR shifts of lithium indenides	50
dopamine transporter binding	50
serotonin transporter binding	50
norepinephrine transporter binding	50
acetylcholinesterase inhibition	50
adenosine activity to guinea pig heart muscle	50
hydroxamic acid inhibition of 5-lipoxygenase	50
cytochrome P450 mediated nitrile toxicity	51
enthalpy of complex formation	52
toxicity of organics	53
toxicity to niterbacter	54
histaminergic activity	55
antineoplastic activity	56
carcinogenicity	57
dihydrofolate reductase inhibitors	58
mutagenicity of quinolines	59
polarizability	60
Hammet parameters	61
vasodilator activity	62
free radical formation	63
metabolism of benzoic acids	64
β-caroline activity	65
genotoxicity	66
GABA-agonist activity and uptake inhibition	67
α_{1a}-adrenoceptor antagonist activity	68

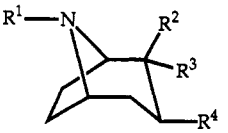
[a] For an additional review of the literature, see ref 24f.

the importance of the polarizability coefficient, i.e., more polarizable cations are better able to bind to polarizable aromatic π systems and thus have a higher (positive) affinity.

Table I provides a list of several diverse properties for which similar comparisons have been made, together with the appropriate references. For the remainder of this Account, we prefer to focus on recent developments which illustrate the current frontiers in this technology.

Condensed Phase Effects. Since many biological properties are measured by in vivo assays, there is a possibility that theoretical descriptors calculated using the standard theoretical assumptions of an isolated molecule in the gas phase may be inadequate for correlating condensed-phase properties. We and others have begun recently to explore the use of theoretical descriptors calculated by an extension of the AM1 Hamiltonian to include solvation effects. The model, known as AM1-SM2,[43] includes self-consistently both electronic polarization and local first-solvation shell effects; the formalism has been reviewed.[43b] Such a

(43) (a) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213. (b) Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 629.



R^1 = H, CH_3 R^2 = H, CO_2H, CO_2Me

R^3 = H, CO_2Me R^4 = HO, BzO, Ph, 4-FPh, 4-NH_2Ph, 4-MePhCO_2

Figure 2. Cocaine analogs exhibiting varying degrees of inhibition of dopamine transporter uptake.

calculation leads to a modified TLSER,

$$\log(\gamma) = c_0 + c_1 V'_{mc} + c_2 \Delta G_S^\circ + c_3 \epsilon'_\alpha + c_4 \epsilon'_\beta + c_5 q'^{+} + c_6 q'^{-} \quad (6)$$

where the polarization term, π^*, has been replaced with the calculated free energy of solvation, ΔG_S°, and the prime (′) on the remaining descriptors indicates that they are calculated under the influence of solvation. If the substrate interactions governing the property of interest take place either in solution or in a highly polar active site, it is expected that the solvated charges and MO energies will deliver better correlations. Additionally, for properties dominated by transport phenomena, e.g., a toxicity assay which is impacted by the speed with which a substrate passes through a hydrophobic membrane separating two aqueous compartments, the ΔG_S°(aq) descriptor should be important.

Thus, aqueous solvation appears to play a role in the observed[44] inhibition of dopamine[45] transporter uptake exhibited by the cocaine analogs illustrated in Figure 2. While the AM1 gas-phase descriptors in regression equation 4 are unsuccessful in producing a statistically relevant correlation, the AM1-SM2 descriptors of eq 6 predict

$$\log \frac{1}{K_i} = 7.451 - \frac{1.622}{10}\Delta G_S^\circ(\text{aq}) - 49.000 q'^{+} \quad (7)$$

$N = 13$ $R = 0.876$ $s = 0.675$ $F = 16.6$

In another regression, aimed at explaining the affinity of a series of 3-benzazepine cations for the dopamine D_1 receptor, Alkorta, Villar, and Perez observed a similarly important dependence on the free energy of solvation,[46]

$$\log \frac{1}{K_i} = 4.291 + \frac{0.810}{10}\Delta G_S^\circ(\text{aq}) - 0.655\mu^* \quad (8)$$

$N = 13$ $R = 0.853$ $s = 0.367$ $F = 13.4$

where μ^* is the coordinate of the molecular dipole

(44) Carroll, F. I.; Lewin, A. H.; Boja, J. W.; Kuhar, M. J. *J. Med. Chem.* **1992**, *35*, 969.

(45) For a study of the effect of solvation on dopamine structure, see: Urban, J. J.; Cramer, C. J.; Famini, G. R. *J. Am. Chem. Soc.* **1992**, *114*, 8226.

(46) (a) Alkorta, I.; Villar, H. O.; Perez, J. J. *J. Comput. Chem.* **1993**, *14*, 620. (b) These authors employed the earlier developed AM1-SM1 model described in the following: Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305, 9901. (c) We have removed a polarizability descriptor of questionable statistical relevance from the regression of Alkorta et al. to facilitate comparison of eqs 7 and 8.

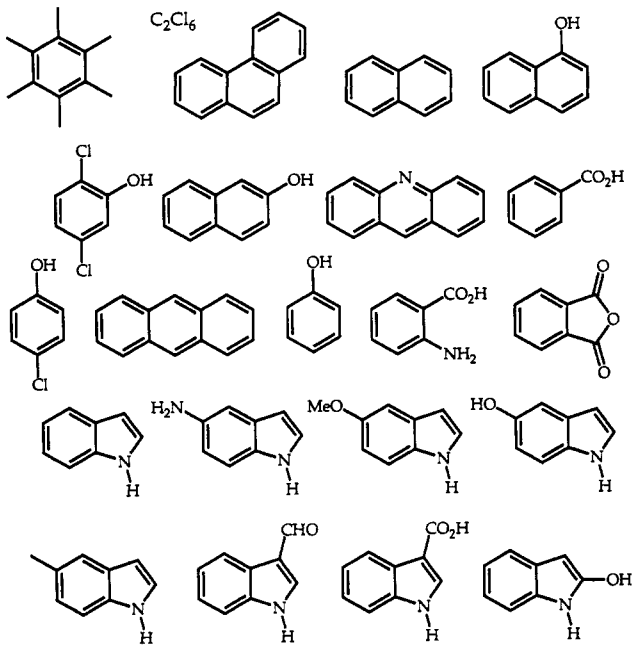


Figure 3. Compounds whose solubility in supercritical carbon dioxide was used in developing regression equations 9 and 10.

moment orthogonal to an activity-weighted average of all substrate dipole moment vectors.

Equations 7 and 8 are statistically on the borderline of being useful in a quantitative sense, but provide important qualitative information about the processes under study. Thus, not only are solvation free energies required to obtain any kind of correlation, but for the cocaine analogs it is only the solvation energies of the neutral molecules which are useful, while for the 3-benzazepines, only the cation solvations serve. In the latter instance, this suggests that it is the cation which interacts with the receptor, and the sign of the coefficient in eq 8 indicates that higher costs of desolvation decrease affinity for the receptor, as would be expected for cation–π interactions such as those mentioned earlier. In eq 7, the negative sign of the solvation free energy coefficient indicates that better solvated substrates are more active, suggesting that the binding sight is rich with hydrogen bonds and/or electrostatic interactions, and substrate–ligand interactions resemble aqueous solvation but without the entropic cost to the solvent.

Property Extrapolation. Finally, we provide an example of this formalism being used in a predictive mode in order to avoid (for as long as possible) dangerous and costly experiments. Thus, the military has some interest in discovering practical ways to decontaminate items which have been exposed to chemical weapons agents, e.g., sulfur or amine mustards or phosphonate-derived nerve agents. Moreover, in order to test emerging decontamination technologies, weapons-agent simulants are required which display many of the physicochemical features of the agent, with the important exception of toxicity. One such technology is supercritical fluid extraction, in particular with carbon dioxide as the extractant. Thus, we have correlated the supercritical-CO_2 solubility at 308 K of the series of compounds illustrated in Figure 3 with MNDO-derived TLSER descriptors to obtain[47]

(47) Famini, G. R.; Wilson, L. Y. *J. Phys. Org. Chem.*, in press.

$$\log S_{CO_2}^{14\ MPa} = -8.370 - \frac{0.604}{0.1}\pi^* + \frac{104.4}{10}\epsilon\beta + 24.350q^+ - 22.098q^- \quad (9)$$

$N = 19$ $R = 0.928$ $s = 0.477$ $F = 22$

and

$$\log S_{CO_2}^{20\ MPa} = -18.083 - \frac{0.437}{0.1}\pi^* + \frac{156.7}{10}\epsilon\beta + 23.207q^+ - 20.783q^- \quad (10)$$

$N = 18$ $R = 0.939$ $s = 0.431$ $F = 24$

where the two equations refer to pressures of 14 and 20 MPa, respectively. Analysis of descriptor t-scores indicates the electrostatic terms q^+ and q^- to be accounting for the majority of the correlation; the signs of the coefficients are consistent with viewing the solute–solvent interactions as being dominated by "hard" type interactions. In addition, the negative sign of the polarizability coefficient may be interpreted as illustrating that weakened van der Waals forces in the crystalline solute enhance solubility in the supercritical fluid. Similar observations have been made with regard to this last point by Politzer et al., who employed a different theoretically based regression using descriptors for molecular volume and electrostatic variance and balance.[48]

With regression 9 in hand, we may easily calculate the appropriate descriptors for various chemical weapons agents and predict their solubility under these

(48) (a) Politzer, P.; Murray, J. S.; Lane, P.; Brinck, T. *J. Phys. Chem.* **1993**, *97*, 729. (b) Murray, J. S.; Lane, P.; Brinck, T.; Politzer, P. *J. Phys. Chem.* **1993**, *97*, 5144. (c) Politzer, P.; Murray, J. S.; Concha, M. C.; Brinck, T. *J. Mol. Struct. (THEOCHEM)* **1993**, *100*, 107.
(49) Famini, G. R.; Marquez, B. C.; Wilson, L. Y. *J. Chem. Soc., Perkin Trans. 2* **1993**, 773.
(50) Famini, G. R.; Wilson, L. Y. Unpublished observations.
(51) Famini, G. R.; DeVito, S. C. In *Biomarkers of Human Exposure to Pesticides*; Blancato, J., Saleh, M., Eds.; American Chemical Society: Washington, DC, in press.
(52) Raevsky, O. A.; Grigor'ev, V. Y.; Zefirov, N. S. *J. Chim. Phys. Phys.-Chim. Biol.* **1992**, *89*, 1747.
(53) Gao, C.; Govind, R.; Tabak, H. H. *Environ. Toxicol. Chem.* **1992**, *11*, 631.
(54) Tang, N. H.; Blum, D. J. W.; Nirmalakhandan, N.; Speece, R. E. *J. Environ. Eng.* **1992**, *118*, 17.
(55) Luque, F. J.; Saauz, F.; Illas, F.; Pouplana, R.; Smeyers, Y. G. *Eur. J. Med. Chem.* **1988**, *23*, 7.
(56) Lewis, D. F. V.; Ioannides, C.; Parke, D. *Toxicol. Lett.* **1989**, *45*, 1.
(57) Parke, D. V.; Lewis, D. F. V.; Ioannides, C. In *Chemicals in the Environment*; Richardson, R. R., Ed.; Royal Society of Chemistry: London, 1988; p 45.
(58) De Benedetti, P. G. *J. Mol. Struct. (THEOCHEM)* **1992**, *256*, 231.
(59) Debnath, A. K.; de Compadre, R.; Hansch, C. *Mutat. Res.* **1992**, *280*, 55.
(60) Brinck, T.; Murray, J. S.; Politzer, P. *J. Chem. Phys.* **1993**, *98*, 4305.
(61) Murray, J. S.; Brinck, T.; Politzer, P. *J. Mol. Struct. (THEOCHEM)* **1992**, *87*, 271.
(62) Hemken, H. G.; Lehmann, F.; Pedro, A. *Quant. Struct.-Act. Relat.* **1992**, *11*, 332.
(63) Paulina-Blumenfeld, M.; Hansz, M.; Hikichi, N.; Sooppani, A. O. M. *Free Radical Res. Commun.* **1992**, *16*, 207.
(64) Ghauri, F. Y.; Blackedge, C. A.; Glen, R. C.; Sweatman, B. C.; Lindon, J. C.; Beddell, C. R.; Wilson, I. D.; Nicholson, J. K. *Biochem. Pharmacol.* **1992**, *44*, 1935.
(65) Allen, M. S.; LaLoggia, A. J.; Dorn, L. J.; Martin, M. J.; Costantino, G.; Hagen, T. J.; Koehler, K. F.; Skolnick, P.; Cook, J. M. *J. Med. Chem.* **1992**, *35*, 4001.
(66) Debnath, A. K.; Hansch, C. *Environ. Mol. Mutagen.* **1992**, *20*, 140.
(67) Tsuda, M.; Takada, T.; Miyazaki, M.; Uda, Y.; Kuzuhara, S.; Kitaura, K. *J. Mol. Struct. (THEOCHEM)* **1993**, *280*, 261.
(68) De Benedetti, P. G.; Cocchi, M.; Menziana, M. C.; Fanelli, F. *J. Mol. Struct. (THEOCHEM)* **1993**, *280*, 283.

Table II. Predicted Solubilities in Supercritical CO_2 of Chemical Weapons Agents and Potential Agent Simulants[a]

compound	solubility (mole fraction) Politzer et al.[b]	solubility (mole fraction) Famini and Wilson[c]
agents		
HD	2.02×10^{-2}	6.62×10^{-3}
GB	1.85×10^{-11}	1.41×10^{-10}
VX	7.10×10^{-8}	2.24×10^{-13}
simulants		
$ClCH_2(CH_2)_3CH_2Cl$	1.33×10^{-2}	1.78×10^{-1}
$HOCH_2CH_2SCH_2CH_3$	1.64×10^{-3}	3.60×10^{-2}
$CH_3P(O)(OEt)_2$	2.80×10^{-8}	1.59×10^{-12}
4-fluorobenzyl alcohol	5.51×10^{-4}	8.10×10^{-4}
$HOCH_2CH_2N(i\text{-}Pr)_2$	8.53×10^{-6}	4.54×10^{-6}

[a] At 14 MPa and 308 K. [b] Reference 48c. [c] Reference 47.

conditions. Moreover, we may do the same for a series of simulants. While the absolute solubilities will probably only be qualitative, given the disparate functionalities in the regression set and the weapons agents, that is still sufficient to guide our thinking. Moreover, the *relative* solubilities of an agent and simulant are expected to be more trustworthy, since there the disparity in functionality is much smaller. As a rule, this is about all one should expect from what amounts to an *extrapolation* using eq 9, as opposed to an *interpolation*, where we could probably predict the solubility of another aromatic hydrocarbon with a high degree of accuracy.

Thus, Table II provides *qualitative* predictions using both eq 9 and the corresponding regression of Politzer et al.[48c] It is evident that while sulfur mustard (HD) is expected to be reasonably soluble in supercritical CO_2 under these conditions, the phosphonate-derived nerve agents GB and VX are not. Furthermore, hydroxyethyl ethyl sulfide, which is biologically innocuous by comparison to HD, appears to be a potentially useful simultant for future work on this technology. The *quantitative* discrepancies between our results and those of Politzer et al. must await experimental resolution, but that is a secondary issue.

Concluding Remarks

The fundamental principles of linear free energy relationships are undeniably sound when used in an interpolative sense, and considerable research now exists to suggest that they are robust enough to enable a fair degree of extrapolation as well. Given the intuitively meaningful nature of many of the descriptors employed in such multi- or univariate analyses, LFER offers the opportunity to separate properties and reactivities into combinations of well-defined relationships to other properties, like polarity, acidity, basicity, etc. Once such a relationship is established, it may be used in a predictive fashion for other substrates for which all of the descriptors are available.

The use of theoretical descriptors, still associated with fundamental chemical properties equivalent to those measured experimentally, allows an exactly analogous formalism to be employed. However, the difficulties associated with experimental measurement (assuming that the substrate has even succumbed to synthesis) are avoided. Instead, computer calculation permits economical prediction of all of the required descriptors. Moreover, once the calculation has been performed and the wavefunction is in hand, any number of theoretical descriptors may be derived therefrom; this permits far more regression analyses to be performed than would be possible if every descriptor required some experimental measurement.

Continued exploration of the relationships between LFER, TLSER, and more rigorous mathematical techniques such as PCA will continue to refine our abilities to select orthogonal, chemically meaningful descriptors, with the goal being to span as much of the infinite dimensionality of chemical property space as possible.

We owe much to the energy and enthusiasm of our many mentors, collaborators, and students, whose names may be found in the references, and without whom little would have been accomplished. We especially thank Professor Pete Carr for helpful discussions during preparation of this manuscript.