

Linear Solvation Energy Relationships. Local Empirical Rules – or Fundamental Laws of Chemistry? The Dialogue Continues. A Challenge to the Chemometricians*

Mortimer J. Kamlet,[§] Ruth M. Doherty, George R. Famini and Robert W. Taft

Advanced Technology and Research, Inc., 14900 Sweitzer Lane, Laurel, Maryland 20707, Naval Surface Weapons Center, White Oak Laboratory, Silver Spring, Maryland 20910, U.S. Army Chemical Research, Engineering and Development Command, Aberdeen Proving Ground, Maryland 21010 and Department of Chemistry, University of California, Irvine, California 92717, U.S.A.

Kamlet, M. J., Doherty, R. M., Famini, G. R. and Taft, R. W., 1987. Linear Solvation Energy Relationships. Local Empirical Rules – or Fundamental Laws of Chemistry? The Dialogue Continues. A Challenge to the Chemometricians. – *Acta Chem. Scand.*, Ser. B 41: 589–598.

This paper continues a dialogue with S. Wold and M. Sjöström. It is shown that the disagreement indicated in the title devolves to a question of whether a methodology needs to be universal and all-encompassing to avoid being characterized as “local” or that it suffices that it is general and widespread. A further disagreement derives from an unrealistic assessment by Wold and Sjöström of the experimental precision that may be expected from usual chemical measurements. The authors challenge the chemometric community to join with them in an assessment of the precision of principal components analysis versus linear solvation energy relationships for the correlation and prediction of new experimental data. It is suggested that usual chemometric performance does not match advertised claims.

For several years we have been involved with the chemometricians S. Wold and M. Sjöström (SW/MS) of Umeå University in a disputation regarding the comparative merits of the methodologies of principal components (PC) analysis versus linear solvation energy relationships (LSER) for the correlation, rationalization and prediction of properties that depend on solute–solvent interactions. To describe the current status of this dialogue, without imposing our own bias or interpretation, we repeat verbatim the abstracts of the earlier papers. Where we quote SW/MS, the text is given in italic print.

The dialogue began when SW/MS, correctly identifying LSER as a subclass of linear free energy relationships (LFER), published a paper entitled “Linear Free Energy Relationships. Local

Empirical Rules – Or Fundamental Laws of Chemistry?”¹ The abstract of that paper read: “*Two fundamentally different interpretations of linear free energy relationships (LFERs) and the causes of breakdowns of one-term LFERs to more complex ones have been forwarded: (a) the classical interpretation of LFERs expressing combinations of ‘fundamental’ effects, or (b) an interpretation where LFERs are looked at as empirical models of similarity.*

In this review we argue against the classical interpretation (a). Instead we provide support for the second alternative (b), where LFERs are seen as locally valid linearizations of complicated relationships. The major argument is that the latter interpretation is scientifically preferable since it results in better predictions of new experimental facts. This is illustrated with data from organic reactivity and solvent effect studies.” The solvent effect studies characterized in this manner involved our analysis by the solvatochromic comparison method of solvent effects on UV/visible

*Part 4 in the series “Studies at the Chemometrics – Physical Organic Chemistry Interface”. Ref. 6 is Part 3.

[§]To whom correspondence should be addressed.

spectra of some 4-nitroaniline and 4-nitrophenol derivatives.^{2,3}

Our response was in a paper entitled "Linear Solvation Energy Relationships. Local Empirical Rules – Or Fundamental Laws of Chemistry? A Reply to the Chemometricians".⁴ The abstract of that paper read, in part, as follows: "...In reply to the chemometricians, it is shown that hundreds of physicochemical properties and reactivity parameters of many diverse types are described by equations of the following forms. For solubility properties of multiple solutes in single solvents, or distributions between pairs of solvents,

$$XYZ = (XYZ)_0 + mV_2/100 + s\pi_2^* + a(\alpha_m) + b(\beta_m), \quad (1)$$

and for effects of different solvents on single indicators or combinations of reactants

$$XYZ = (XYZ)_0 + h(\delta_H^2)_1 + s\pi_1^* + a\alpha_1 + b\beta_1, \quad (2)$$

where XYZ is a property or quantity that depends on a solute-solvent interaction, V is a measure of solute molar volume, δ_H is the solvent Hildebrand solubility parameter, and π^* , α and β are the solvatochromic parameters that scale solute and solvent dipolarity/polarizability, hydrogen bond donor acidity, and hydrogen bond acceptor basicity. Evidence is offered that, if the chemometricians are correct, the term *local* must be stretched to include every area in chemistry and many in biology, where physicochemical, biological, toxicological and pharmacological properties depend on interactions between solutes and solvents".

In addition to the terms in the above equations, the LSER methodology sometimes requires a "polarizability correction" parameter δ in combination with π^* , and for certain "family dependent" properties, a coordinate covalency parameter ξ in combination with β . The subscript 1 in the above equations indicates that the parameter applies to the solvent, the subscript 2 indicates that the parameter applies to the solute, and the subscript m indicates that, for compounds which are capable of self-association by amphi-hydrogen bonding, the parameter applies to the non-self-associated "monomer" form.

We also pointed out in that paper⁴ that "if the chemometricians believe that improved predictions are the main purposes of LFER and LSER,

they fail to understand the purpose or the main utility of such correlations, which is in improved *understandings* of the underlying phenomenology. If one compares their contributions to *better understandings*, the substituent parameters in LFERs and the solvatochromic parameters in LSERs convey far more information than Sjöström and Wold's first, second, third, fourth, and n^{th} principal components. Thus, $s\pi^*$ always stands for solvent (or solute) polarity, mV for molar volume, α for hydrogen bond donor (HBD) acidity, and β for hydrogen bond acceptor (HBA) basicity. By way of contrast, θ_1 , θ_2 , θ_3 , and θ_4 bring no physical property to mind" (and, indeed, correspond to different combinations or properties in each different correlation).

SW/MS's reply to the above was given in a paper entitled, "Linear Free Energy Relationships. Local Empirical Rules – Or Fundamental Laws of Chemistry? A Reply to Kamlet and Taft."⁵ The abstract of that paper was as follows: "*In a recent article, Kamlet and Taft challenged our view that linear free energy relationships (LFER), and in particular linear solvation energy relationships (LSER) are best seen as locally valid linearizations of complicated functional relationships. We here give evidence that our view is supported by observed data, while the one of Kamlet and Taft is not.*

We do not challenge the statement of Kamlet and Taft that their LSER fits a large number of solvent effect data with some statistical significance. We show, however, that in these data, as well as other data measured on series of chemical reactions, there are systematic parts which cannot be predicted by general models like the LSER of Kamlet and Taft.

These system specific regularities are often not negligible and sometimes substantial, which means with our vocabulary that a given LFER or LSER is not generally applicable. Indeed, this has often been found by Kamlet and Taft, who frequently discover 'outliers' and unpredictable differences in behavior between families of solvents when applying their LSER.

To be predictively efficient, a model should be recalibrated for each new investigated system. This is easily done with modern data analytic methods, such as PLS modeling." With the above as prologue, we now continue the dialogue by responding to this most recent contribution by SW/MS.

Do LSER or PC analyses lead to better understandings?

In response to our claim that LSERs lead to improved understandings compared with principal components analyses, SW/MS stated as follows: "*Kamlet and Taft state that their 'general' LSER is preferable to statistically estimated local LSERs because the 'general' model gives an understanding of the system while local models in the form of PC models do not. To this we would like to give an answer on two levels.*

First, since locally estimated models give better fit and better predictions, the 'general' model is in some fundamental way incomplete and the 'understanding' thereby reached is not, in our view, a good one. To not mislead users of LFERs and LSERs we have preferred to use locally estimated parameters which are not transportable to other systems.

Second, however, we realize man's desire to use generally applicable concepts and models. If we with understanding mean seeing the relations between the presently investigated system and other ones, we agree that the PC models and their score vectors \mathbf{t} with loadings \mathbf{p} do not immediately convey this understanding. This 'understanding' can, however, be reached by a subsequent rotation of PC solution to agree as well as possible with established scales, so-called target rotation."

We have pointed out elsewhere⁶ a number of illustrative examples of "target rotation" gone awry. For example, Cramer⁷ has indicated that the C principal component in his BC(DEF) principal component set (which we consider the most generally successful of the various chemometrically derived solvent parameter scales) can be related by target rotation to molecular cohesiveness. He has also demonstrated that the C principal component is the leading term influencing solvent effects on UV/visible $E_T(30)$ values. However, for a UV/visible spectral shift to depend on solvent "cohesiveness", there would need to be a change in volume between states, which is a clear violation of the Franck-Condon principle. The reason for Cramer's misconception lay in the fact that, *for that data set*, there was a coincidental covariance between solvent self-association by amphihydrogen bonding and solvent HBD acidity.

A more general and more important criticism of "target rotation" as a means of relating princi-

pal components to "established scales" which quantify (or attempt to quantify) fundamental physicochemical properties of molecules is as follows. The chemometricians have usually been successful in relating their first principal components to properties regarded as fundamental by most physical organic chemists (i.e. molar volume, dipole moment, polarizability functions based on refractive index, dielectric constant, etc.).

However, principal components analysis requires that the second principal component be orthogonal to the first, the third orthogonal to the first and second, and the n^{th} orthogonal to the first through $(n-1)^{\text{th}}$ in n -dimensional space. Since fundamental properties in nature are rarely orthogonal, it logically follows that second, third and n^{th} principal components can very seldom be related to single properties or scales, but rather to combinations of such properties. Except in the hands of skilled practitioners of the art, unravelling the contributions of these combinations or properties to n^{th} principal components can be a formidable, perhaps impossible task.

A recent analysis of the chemometrics of solvent basicity by Maria, Gal, de Franceschi and Fargin⁸ has shown how powerful a tool principal components analysis and target rotation can be in the hands of scientists who have sophisticated *a priori* knowledge of and experience with the property being studied. More frequently, however, "target rotation" has led only to murky generalizations regarding the physical significance of the second through n^{th} principal components.

Or to better predictions?

An example in the latter category, which is of particular interest to us because we have recently examined the same property using the methodology of LSER, is a recent statistical analysis of the partition coefficient by Dunn, Grigoras and Johansson.¹⁰ The study was an extension of earlier work by Dunn and Wold,¹⁰ and in both instances the data were those assembled by Hansch and Leo.¹¹

Dunn and his coworkers set up a log P matrix for 51 solutes for which solvent/water partition coefficients were available for the solvents: 1-octanol, diethyl ether, chloroform, benzene, carbon tetrachloride and hexane. They reported as

follows: "The application of PCFA to the ... data resulted in a 2-component ... model being significant. The first and major component accounted for 77% of the variation of the log *P* data while the second accounts for an additional 17% and a total of 94% of the variance about the mean of the columns of the data matrix being accounted for (corresponding to a correlation coefficient of about 0.97). These results are consistent with our earlier report (i.e. that with Wold)¹⁰ and substantiate the stability of the analysis."

The first PC was found to correlate roughly with solvent-accessible surface area, SASA, and somewhat better ($r = \text{ca. } 0.90$) with solvent-accessible non-polar surface area (SASA minus hydrogen-bonded surface area). No firm conclusion was reached regarding the significance of the sec-

be discussed in detail in future publications. V_1 is intrinsic (van der Waals) molar volume. These equations convey significant information regarding solvent/water partition. They show that the leading terms influencing partition are the cavity and hydrogen bonding terms, with dipolarity and polarizability exerting only a secondary effect. As will be discussed in future papers, the relative cavity terms ($mV_1/100$) are consistent with relative solvent cohesiveness (as measured by δ_{H}^2). The responses to solute HBA basicity are consistent with relative solvent HBD acidity, and the responses to solute HBD acidity are consistent with relative solvent HBA basicity. Moreover, as we have demonstrated for aqueous solubility¹⁵ eqns. 3–6 allow us to dissect free energies of transfer for each solute from each solvent into

$$\begin{aligned} \log K(\text{octanol}/\text{H}_2\text{O}) &= (0.31 \pm 0.4) + (5.35 \pm 0.05)V_1/100 - (1.05 \pm 0.04)\pi^* \\ &+ (0.35 \pm 0.03)\delta - (3.84 \pm 0.06)\beta + (0.10 \pm 0.04)\alpha_{\text{m}} \end{aligned} \quad (3)$$

$n = 245, r = 0.9959, \text{ s.d.} = 0.131.$

$$\begin{aligned} \log K(\text{C}_6\text{H}_6/\text{H}_2\text{O}) &= (-0.01 \pm 0.13) + (6.23 \pm 0.18)V_1/100 - (0.18 \pm 0.12)\pi^* \\ &+ (0.24 \pm 0.08)\delta - (4.79 \pm 0.15)\beta - (2.73 \pm 0.08)\alpha_{\text{m}} \end{aligned} \quad (4)$$

$n = 43, r = 0.9959, \text{ s.d.} = 0.124.$

$$\begin{aligned} \log K(\text{CHCl}_3/\text{H}_2\text{O}) &= (0.01 \pm 0.13) + (6.12 \pm 0.20)V_1/100 + (0.28 \pm 0.17)\pi^* \\ &- (0.06 \pm 0.09)\delta - (3.62 \pm 0.18)\beta - (3.27 \pm 0.09)\alpha_{\text{m}} \end{aligned} \quad (5)$$

$n = 48, r = 0.9944, \text{ s.d.} = 0.156.$

$$\begin{aligned} \log K(\text{CCl}_4/\text{H}_2\text{O}) &= (0.45 \pm 0.14) + (6.25 \pm 0.15)V_1/100 - (0.75 \pm 0.12)\pi^* \\ &- (0.11 \pm 0.06)\delta - (5.28 \pm 0.16)\beta - (3.18 \pm 0.09)\alpha_{\text{m}} \end{aligned} \quad (6)$$

$n = 54, r = 0.9949, \text{ s.d.} = 0.142.$

ond PC. They observed that there was some apparent relationship between it and solute structure; it was large and negative for acidic solutes, near zero for neutral solutes, and the basic solutes were at the other extreme to the acids. No inkling was given as to how these two PCs could be estimated for new solutes in order to achieve predictive ability.

By way of contrast, we had available to us the same data base¹¹ as was available to Dunn and Wold, and, using the methodology of LSER with eqn. 1, we arrived at eqns. (3)–(6) for solvent/water partition coefficients (which extend earlier work by ourselves^{12,13} and Leahy¹⁴) and which will

water into the contributing solute-solvent interaction terms.

Further, parameter estimation rules have been devised which allow the V_1 values and the solvatochromic parameters for potentially thousands of new solutes to be calculated. From these and eqns. 3–6 it becomes possible to predict solvent/water partition coefficients with a precision that is well beyond the "level of exhaustive fit" (*vide infra*).

The above comparison does not misrepresent the relative state-of-the-art in PC analyses vs. LSER. In future publications we shall bring many similar comparisons to the attention of the che-

mometric and physical organic chemical communities. We hope that, in the absence of evidence at least as compelling as that presented here and in these future papers, SW/MS will desist in their claim that PC analysis leads either to better understandings or to more accurate prediction of new experimental data than LFER and LSER.

How shall we define "local"?

To avoid being characterized as "local", need a methodology be "universal and all encompassing" or should "widely applicable" suffice? In their recent response,⁵ SW/MS "do not challenge the statement of Kamlet and Taft that their LSER fits a large number of solvent effect data with some statistical significance", but emphasize the following points: (a) there are outlier points which we exclude from our correlations; (b) many of our correlations are for subsets of the data, e.g. for results in a set of "select" solvents for which π^* values are very nearly proportional to molecular dipole moments; and (c) in some instances it has been necessary to enlist additional solute and solvent properties to those in eqns. 1 and 2, the example cited by SW/MS being when "Kupfer and Abraham find it necessary to take into account such additional effects as may be operative for the XYZ studied, such as solvent viscosity in fluorescence relaxation procedures."

These statements are true, but they are also perfectly consistent with the methodologies of both LSER and PC analysis. The exclusion of a point identified as a outlier is a valid technique for handling data; it is the criteria for identification of a point as such that can result in disagreement. A point may be identified unequivocally as an outlier if it is demonstrably the result of an erroneous measurement or a typographical error in a published report. The problem arises in the demonstration of the error. It is not always possible to prove that an error in measurement or reporting has been made, but evidence can be marshalled to support such a conclusion. For example, the deviation of a single point in a homologous series from a progression of values that is observed for all other members of the series is often indicative of an erroneous data point.

Another legitimate reason for excluding a data point is the failure of that point to conform to the

recognized limits of the property being modeled, whether the treatment is by LSER or by PC analysis. One should not, for example, be able to include in a correlation of non-specific narcotic effects a compound that is known to act by a specific "reactive" mechanism, such as enzyme inhibition. Indeed, a fair measure of the relative efficacy of treatments by the LSER and PC methods could well be their ability to recognize and identify the compounds that should properly be excluded as "outliers".

Further, an important aspect of eqns. like (1) and (2) is that they often provide more information about the compounds that do not fit the correlations than about the ones that do. As an example, we have offered evidence¹⁶ that carboxylic ester "outliers" relative to an equation for toxicity to the golden orfe fish undergo *in vivo* hydrolysis (differences between observed and predicted toxicities were linear with alkaline hydrolysis rates). In a similar vein, identification of outliers to an equation for solubility of organic non-electrolytes in blood led to the finding that less soluble compounds undergo sequestration in hydrophobic pockets in hemoglobin.¹⁷

The identification of compounds which should properly be "outliers" can be impaired by a failure to include as descriptive variables parameters which correctly identify a compound as unique in some molecular property (such as hydrogen bond donor ability). This can lead to serious "poisoning" of the data matrix, as we have suggested¹⁸ has been the case for acetonitrile solvent in a study of solvent effects on indenyllithium ¹³C NMR shifts by Eliasson, Wold and coworkers.¹⁹

The second and third criticisms offered by SW/MS are, in some respects, two sides of the same coin. In an example of the former case, we have used an additional parameter in our ($\pi^* + d\delta$) formalism to accommodate classes of compounds that are distinguishable from one another in a specific aspect, viz. polarizability.²⁰ The δ parameter is quantized and has discrete values of 0.0 for non-polychlorinated aliphatic compounds, 0.5 for polychlorinated aliphatics, and 1.0 for aromatic compounds, reflecting the fact that polarizability characteristics within each class are reasonably similar. Our treatments of subsets have usually been the restriction of analyses to compounds for which the δ parameter is the same (e.g., $\delta = 0.0$ for "select" solvents), and any dependence on it cannot be isolated.

This would be akin to selecting compounds that are observed to cluster together in a plot of one principal component versus another, and treating them separately rather than adding another θ to account for the clustering. In PC analysis there is nothing to be gained by this selection procedure; the principal components are already divorced from physically recognizable properties, so that interpretation is not necessarily clarified by handling subsets of the data. This is not true for LSERs, where great importance is attached to the use of physically meaningful parameters. As an example where we have restricted our correlations to "select" solvents, isolating the solvent dipolarity effect ($s\pi^*$) on Menschutkin reaction rates in the absence of the polarizability effect ($d\delta$) has allowed us to estimate the transition state dipole moments.²¹

As to criticism (c) of SW/MS, we find it difficult to understand why they should object to the use of the appropriate number of variables required to describe all the factors that influence a property such as a fluorescence lifetime, particularly if each variable is explicitly related to a specific type of solute-solvent interaction. In PC analysis, one uses as many principal components as are needed to achieve the desired goodness of fit. In LSER, we operate under the restriction that an additional variable is not used, even one that significantly improves the goodness of fit, unless the property studied is explicitly related to the solute-solvent interaction measured by that additional parameter. The fact that all solution-dependent properties are not influenced by solvent viscosity, for example, should not preclude its use when it clearly influences the XYZ in eqn. 2.

Criticism (c) appears to reflect SW/MS's view that any equation or methodology that falls short of predicting all solute-solvent interactions for all solutes in all solvents with a single *a priori* set of fully defined and restricted parameters must be characterized as "local".

We, on the other hand, prefer the usual dictionary definition of "local" as that which is not *general* or *widespread*. We feel that to characterize as "local" equations, and solvent, solute and substituent parameters which have found such widespread applicability in so many areas of chemistry and biology goes against this definition and should therefore be acceptable neither to

lexicographers nor to chemists, nor even to chemometricians. Our concluding remarks in a recent review article²² which summarize our arguments in support of LSER were as follows:

"The point that we have tried to make in the foregoing discussion is not simply that we have an equation and methodology that allow us to predict solubilities in blood and toxicities to the golden orfe. It is rather that the same β parameter (for non-self-associating compounds) that is the leading term influencing $S(\text{blood})$ and LC_{50} also linearly correlates ^{19}F NMR shifts of 5-fluorindole and ^1H NMR shifts of 2-methylbut-1-ene-3-yne with correlation coefficients of 0.992 and 0.996. The same π^* parameter as will provide important information regarding the mechanism of tadpole narcosis, toxicity to *Photobacterium phosphoreum*, and possibly the 'rapture of the deep' experienced by deep sea divers, also correlates non-hydrogen bonding solvent effects on UV/visible spectra of 47 indicators, and relates twelve other solvent property scales to one another and to molecular dipole moments".

"All of these results fit together in a mutually consistent and mutually supporting framework of biochemical and physicochemical properties and reactivity parameters which validate both the data base and the methodology. That a β value of 0.48 allows equally well the prediction of both the EC_{50} value of acetone solute in the Microtox test and the v_{max} value of 4-nitroaniline indicator in acetone solvent indicates that this β value is likely to be correct, and that it reflects a fundamental physicochemical property of the molecule".

Usual experimental precision and "exhaustive fit"

We next address the important question of what levels of experimental uncertainties chemists and chemometricians expect in usual types of physicochemical measurements. In a section entitled, "*The difference between significant and exhaustive fit*", SW/MS write:⁵

"Another way to express the difference in views between *K-T* and ourselves is that the former regard a model applicable when it fits the data set better than chance (statistically significant fit). All their arguments ... are based on the fact that their LSER shows a statistically significant fit to a large number of data series. This fact we do not dispute.

Their model shows a statistically significant fit to all of these data.

Chemical data have good precision, usually 2–3%. To be of any practical or theoretical interest, a model must fit data better than just on the level of statistical significance. The optimal level of fit where the model explains all regularities in the data we call the exhaustive level. We mean that K–T, to show that their model is generally applicable, should show that their model reaches at least nearly this level of regularity exhaustion.

K–T make no efforts to show how near the level of exhaustion their model comes. In fact their own results indicate that the fit of their LSER to measured data is significantly less than exhaustive. In Table I we tabulate the residual standard deviations as given in Ref. 1 (our Ref. 4) for the ten data sets (a through k) they use as illustrations of their LSER applicability.”

We see that in all cases where experimental precision is known, the fit of the K–T LSER is not close to this precision. In several cases the fit is much worse (examples a, c, d, e).

Let us now consider how real-world levels of experimental precision compare with usual LSER standard deviations. In example a, a correlation of molar solubilities in water of 93 liquid aliphatic solutes,²³ we reported an r value of 0.994 and an s.d. of 0.144 log units (we have since expanded the correlation to include 105 solutes, with $r = 0.995$ and s.d. = 0.137 log units).²⁴ SW/MS state that the experimental precision of such measurements is 0.03 log units. The data that we used were from a large number of different laboratories, and to evaluate the usual extent of agreement between presumably replicate measurements of solubilities by different researchers, as well as the vulnerability of the calculational method to a datum which is available from only one source (as was the case for many of our data), some log molar solubilities in water at 25°C assembled from different sources by Yalkowsky and coworkers²⁵ are illustrative. Ranges for a number of representative solutes are as follows:

CCl ₄	–1.99 to –2.30
CH ₃ CH ₂ NO ₂	–0.22 to –0.44
CH ₃ COOCH ₃	+0.63 to –0.18
C ₆ H ₅ NO ₂	–1.80 to –2.09
cyclohexanone	–0.05 to –0.61
C ₄ H ₉ COCH ₃	–0.46 to –1.66
1-C ₄ H ₉ OH	+0.08 to –0.05.

These are not extreme examples, but are fully representative of the current state-of-the-art in aqueous solubility measurements.

As another example, Tewari and coworkers²⁶ recently compared their newly determined aqueous solubilities with literature results. Some representative comparisons of log S_w values are as follows:

C ₆ H ₅ CH ₃	–2.20 vs. –2.23
1,4-C ₆ H ₄ (CH ₃) ₂	–2.69 vs. –2.83
1-C ₄ H ₉ Br	–2.20 vs. –2.38
C ₈ H ₁₈	–5.02 vs. –5.22
CH ₃ COOC ₄ H ₉	–1.27 vs. –0.69.

Also, these workers' log S_w value for nitrobenzene is –1.51, to be compared with the range of –1.80 to –2.09 reported by Yalkowsky (above).

Although experimental determinations of aqueous solubility for some classes of compounds may be accurate to within 0.03 log unit, and some techniques may be precise to within these limits, accuracy and precision are not the same, and the limits quoted by SW/MS are certainly not applicable to the collection of data that we treated. The standard errors of the estimates of our solubility, partition and toxicity correlations compare quite favorably with the precision and accuracy of the data available to us. Indeed, based on the definition by SW/MS, and the reproducibility between laboratories shown above, we would consider that an s.d. of 0.30 log units would constitute a reasonable level of exhaustive fit for prediction of aqueous solubility of liquid non-electrolyte solutes. Our standard deviations are well below this value (the reader might also wish to consider how relatively close to exhaustive fit the LSER predictions of solvent/water partition coefficients given by eqns. 3–6 are, compared with the PC analysis predictions by Wold, Dunn and coworkers discussed above).

The next of what SW/MS considered to be particularly bad examples of non-exhaustive fit to LSER involved free energies of transfer between solvents of tetramethylammonium chloride dissociated ions. We reported an s.d. of about 0.6 kcal mol^{–1}; SW/MS estimated that the experimental precision of such measurements should be 0.1–0.2 kcal mol^{–1}. In the original paper²⁷ we had described, as follows, how the free energies of transfer had been determined. “The data to be treated are the free energies of transfer of the

tetramethyl- and tetraethylammonium chloride, bromide and iodide dissociated ions (DI) and ion pairs (IP) from methanol to up to seventeen widely varying aliphatic solvents. Abraham has reported molar solubility data for the tetraalkylammonium halides in these solvents and, *using literature data for ion pair association constants*, has calculated free energies of solution of the dissociated species $R_4N^+ + X^-$ and ion pairs R_4NX .

Knowing the approximations involved in the estimation of ion pair association constants in organic solvents of widely varying polarity, most knowledgeable solution chemists would agree that precisions near $0.6 \text{ kcal mol}^{-1}$ for the free energy of transfer estimates are at or beyond the state-of-the-art for such data. Again there should be no question that here our LSER correlations fully meet all requirements for exhaustive fit.

Another of SW/MS's bad examples involved solvent effects on the UV/visible spectrum of 3,5-dinitroaniline. We reported s.d. = $0.10 \times 10^3 \text{ cm}^{-1}$; SW/MS estimated that the precision of the measurements should be $0.01\text{--}0.03 \times 10^3 \text{ cm}^{-1}$. Here, with modern instrumentation (although modern instrumentation was not used for the data that were correlated), they are correct about the precision that is possible. They are not correct, however, as to what constituted exhaustive fit to UV/visible spectral measurements.

The reason is that, although the numbers being correlated are the positions of the UV/visible band maxima, the property that should fit the LSER is the zero-zero transition energy. Many workers, including most recently Nicolet and Laurence,²⁸ have noted that UV/visible band shapes change from solvent to solvent. This is because of $n \rightarrow \pi^*$ bands lying under the $\pi \rightarrow \pi^*$ bands, band overlap with the tails of higher energy Rydberg bands, and vibrational fine structure. As a consequence, the correspondence of ν_{max} values with the zero-zero transition energies, even for "well behaved" spectra is seldom better than $0.10 \times 10^3 \text{ cm}^{-1}$. On the basis, we again feel that the precision of our correlations of UV/visible spectral data meet SW/MS's requirements for exhaustive fit.

Indeed, we are grateful to SW/MS for introducing us to the concept of exhaustive fit, which we feel validates rather than weakens our LSER methodology. The comparisons confirm that if

"local effects" have, indeed, influenced the properties that we have correlated by eqns. 1 and 2, their contributions to the numerical data that have been reported have usually, but not always, been below the level of "exhaustive significance" (which term we feel should become at least as important to the correlational chemist and the chemometrician as "statistical significance").

The above notwithstanding, we do not feel that any four parameter equation should, unassisted by other input properties, need to meet SW/MS's exhaustive fit criterion to avoid being characterized as "local". We do not disagree that myriad other solute and solvent properties may influence solute-solvent interactions. We have mentioned viscosity and the "polarizability correction" parameter, δ , in the $(\pi^* + d\delta)$ formalism, as well as the coordinate covalency parameter, ξ , which is used to correlate "family dependent" basicity properties. Other additional input parameters that come to mind are $\log P_{\text{atm}}$, which we use in the correlation of gas/water partition coefficients and Henry's law constants,²⁴ and $\log(L^{16})$, the Ostwald coefficient in hexadecane, which we have used in polymer solubility and GLC correlations to measure the energetics of the transfer process from the vapor phase to a condensed phase wherein the solute interacts only by virtue of dispersion forces (i.e. those forces not adequately measured by π^* , α and β).

Also, we have reported²⁹ that, for reasons that are not completely clear, but which may involve π - π interactions (vertical stacking) in the neat aromatic liquid solute, different aqueous solubility relationships are required for aliphatic and aromatic solutes (we feel that energies required to separate the single solute molecules from the bulk liquid solutes may have larger endoergic dependences on π^* in the aromatic than the aliphatic series). If these differences reflect true natural phenomenology, should this make us subject to SW/MS's criticism for treating the aliphatics and aromatics as different subsets of the data? Further, in aqueous solubility correlations, to account for the endoergic process of converting the solid solutes to "supercooled liquids" at 25°C , an additional term in (m.p. -25) needs to be added to eqn. 1. Should this mean that eqn. 1 should be characterized as "local" because it will not accommodate solid solutes in the absence of such a term? We think not.

A challenge to the chemometricians

The claim by SW/MS (and by chemometricians in general) that PC analyses lead to better predictions of new experimental facts than LFER or LSER has too long gone unchallenged. After extensive perusal of the chemometrics literature, we have come to the conclusion that the chemometric product rarely satisfies the claims made in the chemometric advertisements.

"Exhaustive fit" can be a two-edged sword. Let us use this criterion for the comparison of the methodologies. We claim to have surpassed the level of exhaustive fit in the correlation of solvent/water partition coefficients, free energies of transfer of tetraalkylammonium halides between solvents, and solvent effects on UV/visible spectra, as discussed above. Other published correlations where the methodology of LSER has met or surpassed this criterion include aqueous solubilities,^{24,29} large numbers of HPLC correlations,^{30,31} and toxicities of organic non-electrolytes to the golden orfe fish¹⁶ and *Photobacterium phosphoreum*.³² Can the same be said for the chemometric approach? We know of few, if any, examples of achievements of exhaustive fit by PC analysis of any solubility, partition or toxicology property.

We now invite the chemometrics community to analyze these same data bases using their methodology, and to compare their results with ours in terms of both the precision of the correlations and the information that can be gleaned from them regarding the solute-solvent interactions involved. In the light of their continued derogation of our methodology and exaltation of their own, we feel that this is a challenge that SW/MS can hardly ignore. However, we now extend the challenge to other chemometricians. Let us compare the methodologies with any solubility, partition, or toxicology data base of their choice which is amenable to treatment by both methodologies. Let us agree to publish the results in papers simultaneously submitted to the same journal.

We do think it fair, to advise the chemometric community of some preliminary comparisons that we have carried out with data sets which should strongly favor the PC approach over LSER. We have mentioned that we consider Cramer's BC(DEF) principal component set⁷ to be the most generally useful of the chemometrically derived solvent property scales; it is also usually the most

accurate. Cramer's five components were initially derived for 114 compounds from a six-property matrix comprising log (aqueous activity coefficient), log (octanol/water partition coefficient), molar refractivity, boiling point, molar volume, and enthalpy of vaporization. Based on this first set of PC parameters, Cramer then devised an additive-constitutive method for the estimation of new BC(DEF) values, and used these to predict a wide variety of properties, including new aqueous activity coefficients and octanol/water partition coefficients.

As we shall report in detail elsewhere, despite the fact that the solvatochromic parameters were derived mainly from UV/visible, NMR, and IR spectral measurements, whereas the two solubility properties contributed to the formulation of BC(DEF), LSER and BC(DEF) perform about equally well for the prediction of new aqueous activity coefficients and octanol/water partition coefficients. With other solubility properties, such as other solvent/water partition coefficients and HPLC capacity factors, LSER has performed significantly better than BC(DEF) in the examples so far tested. Thus, we have abundant reason to believe that, in addition to providing better understandings, LSER equals or surpasses the predictive accuracy of PC analysis (as currently practiced, if not in theory).

Concluding comments

One further statement by SW/MS requires rebuttal, namely that⁵ "*Since K-T now have about ten adjustable parameters (those in eqns. 1 and 2 plus δ and ζ) in their LSER, their use of multiple regression for the data analysis necessitates 40-50 data points per analyzed series to make the RSD a good measure of the predictive power of the model.*" After so many years of analyzing each other's methodologies, SW/MS should know better. They must certainly be aware that four of the parameters apply only to solvent properties (subscript 1), and four to solute properties (subscript 2), and that for non-self-associating compounds the parameters are usually identical. Further, unlike PC analysis, where one tries as many adjustable parameters or principal components as are required to give the desired fit, we have set forth strict requirements that need to be met before an additional solvatochromic parameter is

used. Thus, where there is no change of volume between states, V_2 drops out of eqn. 1 and $(\delta_{\text{H}}^2)_1$ drops out of eqn. 2. If the solutes are not hydrogen bond donors, the $(\alpha_m)_2$ and β_1 terms drop out, etc. As a result, we have rarely used as many as five parameters in our LSER, and usually use fewer than four.

References

- Sjöström, M. and Wold, S. *Acta Chem. Scand., Ser. B* 35 (1981) 537.
- Kamlet, M. J., Abbod, J.-L. M. and Taft, R. W. *J. Am. Chem. Soc.* 99 (1977) 6027.
- Kamlet, M. J., Jones, M. E., Taft, R. W. and Abboud, J.-L. M. *J. Chem. Soc., Perkin Trans 2* (1979) 342.
- Kamlet, M. J. and Taft, R. W. *Acta Chem. Scand., Ser. B* 39 (1985) 611.
- Wold, S. and Sjöström, M. *Acta Chem. Scand., Ser. B* 40 (1986) 270.
- Kamlet, M. J. and Carr, P. W. *J. Phys. Chem. Submitted for publication.*
- Cramer, R. D., III. *J. Am. Chem. Soc.* 102 (1980) 1837; *Ibid.* 1849.
- Maria, P. C., Gal, J. F., de Francheschi, J. and Fargin, E. *J. Chem. Soc., Perkin Trans. 2* (1987) 483.
- Dunn, W. J., III, Grigoras, S. and Johansson, E. In: Dunn, W. J., III, Bloch, J. H. and Pearlman, R., Eds., *The Partition Coefficient*, Pergamon Press, New York 1986.
- Dunn, W. J. and Wold, S. *Acta Chem. Scand., Ser. B* 32 (1978) 536.
- Leo, A. J. and Hansch, C. *Pomona College Medicinal Chemistry Project Data Bank*, Claremont, California.
- Kamlet, M. J., Abraham, M. H., Doherty, R. M. and Kamlet, M. J. *J. Am. Chem. Soc.* 106 (1984) 464.
- Taft, R. W., Abraham, M. H., Famini, G. R., Doherty, R. M. and Kamlet, M. J. *J. Pharm. Sci.* 74 (1985) 807.
- Leahy, D. *J. Pharm. Sci.* 75 (1986) 629.
- Kamlet, M. J., Doherty, R. M., Abboud, J.-L. M., Abraham, M. H. and Taft, R. W. *J. Pharm. Sci.* 75 (1986) 336.
- Kamlet, M. J., Doherty, R. M., Abraham, D. J., Veith, G. D., Taft, R. W. and Abraham, M. H. *Environ. Sci. Technol.* 22 (1987) 149.
- Kamlet, M. J., Abraham, D. J., Doherty, R. M., Taft, R. W. and Abraham, M. H. *J. Pharm. Sci.* 75 (1986) 350.
- Kamlet, M. J. and Taft, R. W. *Acta Chem. Scand., Ser. B* 40 (1986) 619.
- Eliasson, B., Johnels, D., Wold, S. and Edlund, U. *Acta Chem. Scand., Ser. B* 36 (1982) 155.
- Taft, R. W., Abboud, J.-L. M. and Kamlet, M. J. *J. Am. Chem. Soc.* 103 (1981) 1080.
- Kamlet, M. J., Doherty, R. M., Famini, G. R. and Taft, R. W. *Unpublished results.*
- Kamlet, M. J., Doherty, R. M., Abboud, J.-L. M., Abraham, M. H. and Taft, R. W. *CHEMTECH* 16 (1986) 566.
- Taft, R. W., Abraham, M. H., Doherty, R. M. and Kamlet, M. J. *Nature (London)* 313 (1985) 384.
- Kamlet, M. J., Doherty, R. M., Abboud, J.-L. M., Abraham, M. H. and Taft, R. W. *J. Pharm. Sci.* 75 (1986) 336.
- Valvani, S. C., Yalkowsky, S. H. and Roseman, T. J. *J. Pharm. Sci.* 70 (1981) 502.
- Tewari, Y. B., Miller, M. M., Wasik, S. P. and Martire, D. E. *J. Chem. Eng. Data* 27 (1982) 451.
- Taft, R. W., Abraham, M. H., Doherty, R. M. and Kamlet, M. J. *J. Am. Chem. Soc.* 107 (1985) 3105.
- Nicolet, P. and Laurence, C. *J. Chem. Soc., Perkin Trans. 2* (1986) 1071.
- Kamlet, M. J., Doherty, R. M., Abraham, M. H., Carr, P. W., Doherty, R. F. and Taft, R. W. *J. Phys. Chem.* 91 (1987) 1996.
- Leahy, D. E., Carr, P. W., Pearlman, R. S., Taft, R. W. and Kamlet, M. J. *Chromatographia* 21 (1986) 473.
- Sadek, P. C., Carr, P. W., Doherty, R. M., Kamlet, M. J., Taft, R. W. and Abraham, M. H. *Anal. Chem.* 57 (1985) 2871.
- Kamlet, M. J., Doherty, R. M., Veith, G. D., Taft, R. W. and Abraham, M. H. *Environ. Sci. Technol.* 20 (1986) 690.

Received April 10, 1987.