On the Interpretation of the Solvent-Induced ^{13}C NMR Shifts of Indenyllithium. Arguments against the Approach of Kamlet and Taft

Bertil Eliasson, Dan Johnels, Svante Wold,§ Ulf Edlund and Michael Sjöström

Department of Organic Chemistry, Umeå University, S-901 87 Umeå, Sweden

Eliasson, B., Johnels, D., Wold, S., Edlund, U. and Sjöström, M., 1987. On the Interpretation of the Solvent-Induced ^{13}C NMR Shifts of Indenyllithium. Arguments against the Approach of Kamlet and Taft. – Acta Chem. Scand., Ser. B 41: 291–296.

It is shown that the Procrustean rationalization of solvent effects on ^{13}C NMR shifts of indenyllithium by means of the linear free solvation energy relationship (LSER) of Kamlet and Taft leads to erroneous conclusions. Their claim that the highest charge density in the indenyl anion is located at the quatenary carbons contradicts chemical knowledge. We also reject their arguments that acetonitrile acts as a hydrogen bond donor to the anion, in particular to positions of low π-charge.

In contrast, our analysis of the measured NMR data by multivariate statistical models gives a direct interpretation of the patterns in the data in terms of (a) a solvent effect predictable from other systems by means of, for instance, the Kamlet-Taft scale π^*, and (b) a system-specific solvent effect. These two effects account for about 60 % and 30 % respectively, of the variance in the measured data. This corroborates our view of LSERs and similar models as locally valid linearizations of complicated relationships.

In recent articles in this Journal,[1,2] Kamlet and Taft (henceforth abbreviated K-T) challenge our use of multivariate statistical methods for the analysis of physical organic chemical data.[3–5] In particular, they dislike our conclusion that so-called Linear Free Energy Relationships (LFER) are best interpreted as local linearizations of complicated relationships.

K-T prefer to see LFERs, of which the Linear Solvation Energy Relationship (LSER) they developed is an example, as "natural laws", fundamental models with general scope and applicability. Their basic argument is that the K-T LSER correlates data as diverse as solvent effects on the UV-spectrum of 3,5-dinitroaniline and the toxicity of organic solvents to the Fathead Minnow. Hence, K-T consider their LSER *not* to be a "local" model.

In our reply to K-T in this Journal,[6] we point out that there is no inconsistency between our interpretation of the K-T LSER as a local model and their finding that the LSER correlates a wide range of data. We show that their correlations are not *exhaustive,* i.e. that there are substantial systematic parts in the data that are *not* accounted for by the K-T LSER correlation. The existence of such system-specific regularities is strong support for the local model interpretation and is difficult to explain with the K-T philosophy.

In the second paper,[2] K-T now go into a detailed criticism of a paper of ours in which we analysed solvent effects on the ^{13}C NMR chemical shifts of indenyllithium by means of multivariate statistical models.[7] They claim that we have overlooked the existence of a single hydrogen bond donor in the set of solvents, and that we thereby use "poisoned data" to generate misinformation. They generalize this claim to suggest that our multivariate analysis approach is successful only when the data have been pre-screened by physical organic chemists.

As physical organic chemists, we cannot let this

---

§To whom correspondence should be addressed.

serious criticism stand unanswered. In the present paper we scrutinize their arguments, and find them chemically and statistically incorrect. Our conclusion, therefore, is that rather than demonstrating the general applicability of their LSER, K-T manage to support our view that the uncritical application of a model developed for one type of system to another type of system is likely to result in confusion.

The Kamlet-Taft (K-T) conclusions

To structure the discussion, we first briefly review the conclusions reached by K-T from their analysis of the ^{13}C NMR data of indenyllithium in twelve solvents by means of the K-T LSER with three terms:

$$XYZ = XYZ_0 + s\pi^* + a\alpha + b\beta \qquad (1)$$

Here, XYZ is a measured property (i.e. relative ^{13}C NMR shift) of a solute (indenyllithium) in a series of solvents. The solvents are characterized by the parameters π^*, α and β, which have been derived by K-T for a large number of solvents (see e.g. Ref. 1).

The conclusion of K-T are recited below. The acronyms HBD and HBA stand for hydrogen bond donors and acceptors, respectively. The notation **1** in paragraph (d) refers to the indenyl anion.

(a) "Reflecting strong dependencies on solute HBD acidity, the two-parameter correlations for the C_5–C_6H_{12}, C_1–C_5, and C_2–C_5 chemical shifts deteriorate in quality when the CH_3CN data are included, but goodness of fit is restored when a term in α is also allowed".

(b) "The dependence on solvent π^* values, and hence charge density in the carbanion, is highest at C_{3a} and C_{7a}".

(c) "However, probably because of steric effects, ion pairing and hydrogen bonding at $C_{3a,7a}$ are minimal".

(d) "Hydrogen-bonding by the HBD solvent to **1** exerts its strongest effect at the positions on the five-membered ring".

(e) "Although statistically significant at the 95 % confidence level in only one instance, the small effect of solvent HBA basicity is probably real and most evident at C_1 and C_2".

How the K-T conclusions are wrong and misleading

We shall now look at the K-T conclusions one at a time and scrutinize them with respect to available facts.

(a) Looking at the goodness of fit criteria of K-T in their eqns. 3 to 7 in Ref. 2, we see that *in no case* is the fit chemically or statistically better when the "suspect" solvent CH_3CN is excluded or when the HBD parameter α is included. The statistical F-values for exclusion of the observation are all smaller than 1.0 and the residual standard deviation (SD) increases at most by 0.2, from 0.7 to 0.9 in eqn. 4 and from 0.9 to 1.1 in eqn. 5.

In multivariate analysis, as in all data analysis, it is of utmost importance to find and eliminate outliers in the data before calculating model parameters. Multivariate analysis has the advantage of being able to identify gross outliers directly in the score plot. In the present case, we see (Fig. 1) that CH_3CN does not deviate from the other solvents with π-electrons, but rather that the solvents are divided into two groups, not explicitly modelled by K-T. Looking for moderate outliers, we see in Ref. 7, Table 4 that none of the solvents show significantly large deviations from the multivariate principal components model.

Indeed, when the data are subjected to renewed principal components and PLS analyses, the same results are obtained regardless of whether solvent No. 7, CH_3CN, is included or not (Tables 1 and 2). Thus, the K-T statement is contrary to the facts; the solvent CH_3CN does not poison the data set in the manner claimed.

(b) K-T conclude that the charge density is highest at the quaternary carbons C_{3a} and C_{7a}. This goes against all knowledge concerning the charge distribution in the indenyl anion. MO calculations show that the highest π-charge density is found at positions 1 and 3, followed by position 2 and positions 5 and 6.[8] Alkylation reactions of the indenyl anion, exchange reactions etc. are also consistent with this picture.[9] The high π-charge density at $C_{1,3}$ is also supported by the unique upfield ^{13}C NMR shift of these positions, while the position of the low-field $C_{3a,7a}$ resonance is consistent with a low calculated π-charge at these carbons. The K-T analysis thus leads to the conclusion that those positions in the indenyl anion which according to chemical knowledge

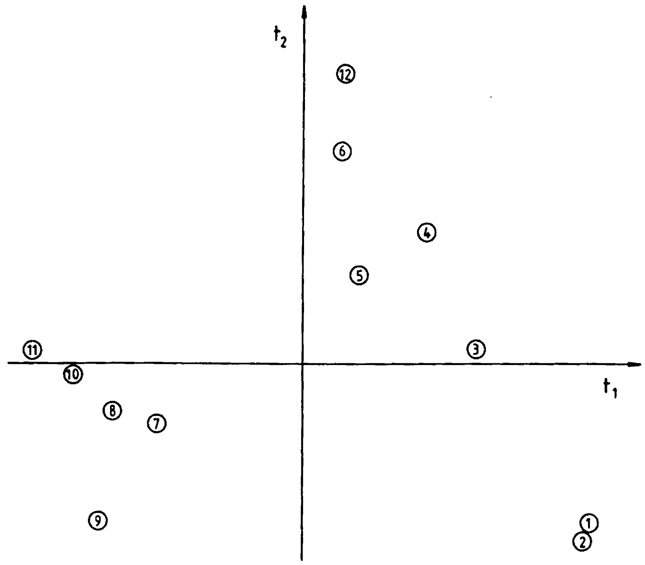

Fig. 1. Plot of first principal component score ($t1 = \theta_1$) against the second ($t2 = \theta_2$) of the ^{13}C NMR data under debate (see Ref. 7). The compounds fall into two groups, with the group to the left containing those with π-electrons. Acetonitrile (No. 7) does not deviate in any respect from this group. The numbering of compounds is shown in Table 1. Anisole has been excluded to make the data agree with those used by K-T. However, anisole is fitted well by the model and falls close to points 1 and 2 in the plot.

have the *lowest* charge density have the ***highest*** charge density.

(c) In spite of their conclusion (b), K-T state that ion pairing and hydrogen bonding is minimal at C_{3a} and C_{7a} because of steric effects. However, a planar carbanion has no orbitals in the molecular plane accessible for overlap with acidic hydrogens. Hence, K-T's steric reasoning is grossly misleading.

(d) Since none of the solvents is a hydrogen bond donor (HBD), and particularly not acetonitrile (see below), this conclusion of K-T is misleading.

(e) K-T believe that hydrogen bonding acceptor basicity probably is real and most evident at

Table 1. Resulting score values (θ_1 and θ_2) for the principal components analyses of the NMR data with and without solvent No. 7 (acetonitrile). The small differences in θ_1 are due to the differences in range for the variables when solvent 7 is included or not. Plots of θ_1 and θ_2 are very similar, however, and the conclusions are not changed by the deletion of acetonitrile from the analysis.

No.	Solvent	θ_1	θ_2	θ_1	θ_2
1	Triethylamine	−9.65	−0.97	−9.22	−0.97
2	iPr-ether	−9.48	−1.01	−9.04	−1.01
3	Et-ether	−5.93	0.03	−5.49	0.03
4	THP	−4.16	0.65	−3.72	0.63
5	DME	−1.94	0.45	−1.50	0.42
6	THF	−1.33	1.13	−0.89	1.11
7	ACN	4.81	−0.38		
8	Pyridine	6.30	−0.35	6.73	−0.41
9	DMSO	6.71	−0.96	7.14	−1.02
10	DMF	7.48	−0.15	7.91	−0.22
11	HMPA	8.79	0.04	9.23	−0.04
12	Dioxane	−1.59	1.52	−1.15	1.49
Residual SD (A = 2)		0.217		0.202	

Table 2. Averages and loadings for the PC-model with two dimensions calculated with and without acetonitrile. Anisole is also excluded to make the data correspond to those used by K-T. The last column in the table indicates the residual standard deviation after each model dimension.

	C–1	C–2	C–3a	C–4	RSD (PCA)
With acetonitrile (without anisole)					
Av.	−21.35	2.37	15.27	5.67	3.29
p–1	0.45	0.60	0.63	0.21	0.52
p–2	−0.57	−0.37	0.73	0.07	0.21
Without acetonitrile and anisole					
Av.	−21.31	2.41	15.44	5.75	3.42
p–1	0.45	0.60	0.62	0.21	0.54
p–2	−0.58	−0.34	0.74	0.04	0.20

C_1 and C_2. In the literature, acetonitrile is commonly considered to be incapable of forming hydrogen bonds to anions.[10] We know of no reports of hydrogen bonds from acetonitrile to systems with basicity similar to that of the indenyl anion. As already discussed, there is no evidence in the present data of any special effect of this solvent. A pK_a difference of 11.2 units between indene (20.1) and acetonitrile (31.3) in DMSO solution has been reported.[11] This implied that hydrogen bonding between acetonitrile and the indenyl anion is most unlikely.

On the use and misuse of multiple regression

A main reason for the erroneous conclusions of K-T is their misuse of multiple regression. As explicitly stated in statistical texts (see e.g. Ref. 12), regression coefficients are difficult or impossible to interpret when the corresponding variables are co-linear. The correlation coefficient for the K-T scales π^* and β for the present data set is 0.92, mainly because all values are positive. Hence the relative magnitudes of the regression coefficients a and b in eqn. 2 depend to a great extent on small changes in the data due to e.g. experimental error.

Moreover, one cannot claim, as K-T do, that two coefficients differ when their confidence intervals overlap. Hence, for example, the coefficients of β in eqns. 8a and 8b in Ref. 2 are *not* statistically different, and their difference is therefore *not* chemically significant.

K-T interpret small differences in correlation coefficients (r) and residual standard deviations (sd) when one point – often that for CH_3CN – is included or deleted as being real. To be interpreted as real, the differences must be statistically significant, which is easily investigated by means of an F-test. As discussed above, the differences on which K-T base their "conclusions" are in *no case* significant in eqns. 3 through 8b in Ref. 2.

Why is the fit of the K-T LSER to the NMR data so bad?

K-T find their model to fit the ^{13}C NMR data of Ref. 7 with an average residual standard deviation of about 0.8 ppm, to be compared with our multivariate model which fits within about 0.2 ppm.

The solvation of lithium salts of delocalized anions has been thoroughly investigated by UV, ESR and NMR spectroscopy during the last 20 years. It is widely accepted that solvent-induced spectral changes for the carbanion mainly reflect lithium solvation or ion pair effects. Under contact ion pair (CIP) conditions in the indenyllithium system, the cation exerts a polarization of the negative charge towards the five-membered ring. In a CIP, the electronic polarization may vary as a result of a change in the *external* cation solvation. An example of this situation is the increase in solvent polarity on going from triethylamine to diethyl ether.

Under solvent-separated ion pair (SSIP) conditions, the ^{13}C NMR chemical shift dependence on the solvent will be less pronounced, since the electric field from the solvated cation will be lower in magnitude. In the intermediate case, when both CIP's and SSIP's co-exist in solution in a fast exchange situation, the ^{13}C NMR spectrum shows only the population-averaged chemical shifts for the CIP and the SSIP. Here, the chemical shift is a sensitive function of the CIP/SSIP equilibrium, where a solvent-induced change in the ion pair structure has a marked effect on the chemical shifts. K-T have overlooked this chemical knowledge and consider ion-pair effects to be only of second-order, hidden in the residuals with SD of 0.8 ppm. Thus, they conclude that our model overfits the data and that our second model dimension "is mainly a scavenger of experimental residuals", which they claim in most instances is statistically insignificant.

If the K-T LSER were a general model for solvent effects, it would, of course, predict satisfactorily the dependence of the ion pair equilibrium constants (in logarithmic form) on the change of solvent by eqn. (1). This would, in turn, lead to a relation between on the one hand the relative amounts of CIP and SSIP, and on the other hand the K-T parameters π^*, α and β. This relation could, for moderate changes in equilibrium constants, be linearized to give eqn. (1).

If now the chemical shifts of each of the ion pairs CIP and SSIP are explained by eqn. (1) and if their relative amounts also obey the same equation, the total result would be that the measured NMR data would be described by eqn. (1) to a precision close to the experimental precision. Hence, the fact that the K-T LSER does not fit the NMR data to a precision even near the experimental precision (about 0.1 ppm) is another indication that their model is *not* of general scope.

The claim of K-T that our second model dimension θ_2 in most cases is statistically insignificant is incorrect. Our significance test is based on cross-validation[13] and shows that predictions of deleted observations are significantly better when the second term is included than when it is not.

Predictions of new facts

K-T find it difficult to see how our multivariate model can be used to make predictions for new systems and molecules.

First, we believe that models are only locally valid (i.e. only locally precise). Hence, for systems dissimilar to the present one, predictions based on our model are just as imprecise as those obtained by the K-T LSER unless a recalibration is made for the new system. This would then result in a new, unrelated model, valid only for the new system (see further below).

Second, within the indenyllithium system and systems similar to it, we need four to five measurements per solvent and per system to obtain estimates of θ_1 and θ_2. This gives simultaneously an estimate of the similarity of a particular solvent to those used in the "calibration phase" (see below).

Local models are thus useful only if predictions beyond the ^{13}C NMR spectra are desired, for instance of proton NMR spectra, IR spectra, solubility, reactivity, ion pairing equilibrium constants etc. A PLS model of the pertinent property and its relation to, say, ^{13}C NMR would then be calibrated against a "training" set of solvents, and could thereafter be used to make predictions of the property in question for new solvents from their ^{13}C NMR spectra.

This *multivariate calibration* approach is described in detail in several papers by ourselves and by others.[14–18] One important experience with this approach is that the predictions are always better for systems similar to those for which the calibration was made than for more dissimilar systems.[18]

Conclusions

After studying the criticism of K-T and the arguments they use as support, we conclude that they have been misled by incorrect use of multiple regression and a failure to understand what we mean by local models.

The main conclusion derived from our multivariate analysis of the solvent-induced ^{13}C NMR shifts of indenyllithium is that there are two "effects", one accounting for about 60 % of the shift variation and which is generalizable to other systems (by, for instance, the use of π^*), and a second "effect" that is system-specific, accounting for about 30 % of the variation in the data.

Secondly, we find that the solvents fall into two groups, one including mainly ethers and the other including solvents with π-electrons (Fig. 1). Hence, one interpretation of the "system-specific" effect would be that it is due to solvent magnetic anisotropy. This grouping of the solvents is overlooked by K-T since they analyse the shifts of the separate atoms individually.

None of our conclusions are altered by the K-T arguments; on the contrary, we feel that they are considerably strengthened.

The chemical conclusions drawn by K-T on the basis of their analysis, i.e. concerning charge distribution, hydrogen bonding of CH_3CN and steric effects, are in considerable conflict with other evidence. It seems that K-T mistakenly believe that the solvent-induced ^{13}C NMR shifts for indenyllithium are due to *intrinsic* solvent carbanion interactions. However, all available evidence indicates that the shifts are due to *changes in the cation solvation,* which induces differences in the charge polarization.

The K-T analysis of the indenyllithium ^{13}C NMR data provides a demonstration of what may happen when extrapolations are made far outside the domain for which a model is developed. The use of a Procrustean methodology,* forcing all observations into a pre-moulded form derived for other systems, may lead to considerable confusion.

Acknowledgements. We are grateful for support by the Swedish Natural Science Research Council (NFR).

---

*After Procrustes, a villain in ancient Greece who fitted his victims into a bed by cutting off parts that stuck out and stretching other parts that were too small.

References

1. Kamlet, M. J. and Taft, R. W. *Acta Chem. Scand., Ser. B 39* (1985) 611.
2. Kamlet, M. J. and Taft, R. W. *Acta Chem. Scand., Ser. B 40* (1986) 619.
3. Wold, S. and Sjöström, M. In: Chapman, N. B. and Shorter, J., Eds., *Correlation Analysis, Recent Advances,* Plenum, London 1978, pp. 1–54.
4. Wold, S., Albano, C., Dunn, W. J., III, Edlund, U., Eliasson, B., Johansson, E., Nordén, B. and Sjöström, M. In: Jöreskog, K. G. and Wold, H., Eds., *Proceedings of The conference on Systems under Indirect Observation,* North-Holland, Amsterdam 1982, pp. 177–207.
5. Sjöström, M. and Wold, S. *Acta Chem. Scand., Ser. B 35* (1981) 537.
6. Wold, S. and Sjöström, M. *Acta Chem. Scand., Ser. B 40* (1986) 270.
7. Eliasson, B., Johnels, D., Wold, S. and Edlund, U. *Acta Chem. Scand., Ser. B 36* (1982) 155.
8. Vos, H. W., MacLean, C. and Velthorst, N. H. *J. Chem. Soc., Faraday Trans. 2, 72* (1976) 63.
9. (a) Meurling, L., *Acta Chem. Scand., Ser. B 28* (1974) 295; *Ibid.* 399; (b) Almy, J. and Cram, D. J. *J. Am. Chem. Soc. 92* (1970) 4316.
10. Popovych, O. and Tomkins, R. P. T. *Nonaqueous Solution Chemistry,* Wiley-Interscience, New York 1981, pp. 62–64.
11. Bordwell, F. G., Drucker, G. E. and Fried, H. E. *J. Org. Chem. 46* (1981) 632.
12. Draper, N. R. and Smith, H. *Applied Regression Analysis, 2nd ed.* Wiley, New York 1981.
13. Wold, S. *Technometrics 20* (1978) 397.
14. Wold, S., Martens, H. and Wold, H. In: Ruhe, A. and Kågström, B., Eds., *Proceedings of the Conference on Matrix Pencils,* Piteå, Sweden, March 1982. *Lecture Notes in Mathematics,* Springer, Heidelberg 1983, pp. 286–93.
15. Lindberg, W., Persson, J.-Å. and Wold, S. *Anal. Chem. 55* (1983) 643.
16. Wold, S., Albano, C., Dunn, W. J., III, Esbensen, K., Hellberg, S., Johansson, E., Lindberg, W. and Sjöström, M. *Analusis (Paris) 12* (1984) 477.
17. Martens, H. and Naes, T. *Trends in Anal. Chem. 3* (1984) 204; *Ibid.* 266.
18. Martens, H. *Multivariate Calibration, Quantitative Interpretation of Non-Selective Chemical Data,* Technical University of Norway, Trondheim 1985. ISBN 82-90394-10-1.

Received November 26, 1986.