Acta Chemica Scandinavica B 39 (1985) 79–91

Screening of Suitable Solvents in Organic Synthesis. Strategies for Solvent Selection

ROLF CARLSON,[a] TORBJÖRN LUNDSTEDT[a] and CHRISTER ALBANO[b]

[a] Department of Organic Chemistry, University of Umeå, S-901 87 Umeå, Sweden and [b] National Swedish Laboratory for Agricultural Chemistry, Box 720, S-901 10 Umeå, Sweden

The problem of selecting test solvents for the study of new organic synthetic methods is discussed. Eight common descriptors of solvent properties for 82 different solvents were analyzed by principal components (PC) analysis. On the basis of the systematic variation of the descriptors, as revealed by the PC-analysis, different strategies for a systematic selection of test solvents are discussed. Among these strategies are selection by a sequential simplex procedure and selection by D-optimal designs.

A problem that is rarely discussed in the vast number of papers published continuously on new synthetic methods is how the chemist has arrived to select the solvent(s) used. Experience, intuition and analogies with other reactions can, of course, often guide the chemist to make an educated guess of a suitable solvent. However, with totally new substrates and/or reagents, prior experimental experience is scarce, and attempts to select solvents by a hypothetical reaction mechanism can lead to mistakes. If the supposed mechanism is false and the reaction does not work in the selected solvents actually studied, it is likely that the chemist will abandon the route and turn to more promising projects. Speculations in unknown reaction mechanisms run the risk of narrowing the perspective of the choice. A more diverse test battery of solvents would account for a variety of solvent "effects" and may allow for a "hit", even if a supposed mechanism should be found to be false later on.

A solvent can play many roles in a reaction: It can interact specifically with species in solution (charge stabilization, ion solvation, hydrogen bonding...). It can serve as a heat transport medium. It can act merely as a diluent, *etc.* It is obvious that any attempt to derive one single, universally valid, model which accounts for all solvent effects in all chemical reactions is doomed to fail. Analysis of how a solvent influences the course of a reaction has been made by numerous different approaches, theoretical as well as purely empirical, (see Ref. 1 for an excellent review). Since nothing is really known with certainty when new synthetic reactions are studied, a theoretical approach to solvent selection is excluded. An empirical approach is therefore necessary. The most common way to empirically analyze solvent effects is to study a rather well defined phenomenon (*e.g.* rates of model reactions, spectroscopic properties, positions of equilibria) in different solvents and then relate changes in the observed phenomenon to changes in known solvent properties. Such an approach can be fruitful when the new reaction bears a strong resemblance to another well

0302-4369/85 $2.50

Table 1. Descriptors and *t*-values in the Principal Components analysis.

Solvent (ordered after descriptor 6, $E_T(30)$)	Descriptors 1 m.p. (°C)	2 b.p. (°C)	3 ε_{25} °C	4 $\mu \cdot 10^{30}$ (C m)	5 n_D^{20}	6 $E_T(30)$ (kcal/mol)	7 $\rho \cdot 10^3$ (kg/m^3)	8 *log P*	*t*-values t_1	t_2
1. Water	0.0	100.0	78.39	6.07	1.333	63.1	0.9982	−1.38	3.07	−1.96
2. Formamide	2.5	210.5	111.0[a]	11.24	1.4475	56.6	1.1134	−1.51	4.11	0.26
3. 1,2-Ethanediol	13	197.3	37.7	7.61	1.4318	56.3	–	−1.93	2.72	−0.25
4. Methanol	−97.7	64.7	32.20	5.67	1.3284	55.5	0.7914	−0.77	1.03	−2.95
5. *N*-Methylformamide	−3.8	180−5	182.4	12.88	1.4319	54.1	1.01	–	5.45	−0.45
6. Diethylene glycol	−6.5	244.8	31.69[a]	7.71	1.4475	53.8	1.109	–	2.26	0.78
7. Triethylene glycol	−4.3	288.0	23.69[a]	9.97	1.4561	53.5	–	–	2.70	1.29
8. 2-Methoxyethanol	−85.1	124.6	16.93	6.81	1.4021	52.3	0.065	0	0.94	−1.20
9. *N*-Methylacetamide	30.6	206.7	191.3[b]	14.65	1.4286[c]	52.0	0.957[d]	−1.05	5.29	0.58
10. Ethanol	−114.1	78.3	24.55	5.77	1.3614	51.9	0.789	−0.31	0.48	−2.51
11. 2-Aminoethanol	10.5	171.0	37.72	7.57	1.4539	51.8	1.018	−1.31	2.16	0.22
12. Acetic acid	16.7	117.9	6.15[a]	5.60	1.3719	51.2	1.0492	−0.17	0.94	−0.67
13. Benzyl alcohol	−15.3	205.5	13.1[a]	5.54	1.5404	50.8	1.042	1.10	0.60	1.53
14. 1-Propanol	−126.2	97.2	20.33	5.54	1.3856	50.7	0.804	0.25	0.12	−2.10
15. 1-Butanol	−88.6	117.7	17.51	5.84	1.3993	50.2	0.8098	0.89	0.12	−1.27
16. 2-Methyl-1-propanol	−108	107.7	17.93	5.97	1.3959	49.0	0.794[f]	0.83	−0.05	−1.56
17. 2-Propanol	−88.0	82.3	19.92	5.54	1.3772	48.6	0.786	0.05	0.18	−1.90
18. 2-Butanol	−114.7	99.6	16.45	5.54	1.3972	47.1	0.8080	0.61	−9.24	−1.69
19. 3-Methyl-1-butanol	−117.2	130.5	14.7	6.07	1.4071	47.0	0.8092	1.16	−0.28	−1.23
20. Cyclohexanol	25.2	161.1	15.0	6.20	1.4548[d]	46.9	0.962	1.23	0.48	1.02
21. 4-Methyl-1,3-dioxol-2-one	−48.8	241.7	65.1	16.7	1.4209	46.6	1.204	–	3.31	0.23
22. 2-Pentanol		119.0	13.82[e]	5.54	1.4064	46.5	0.810	–	−0.03	−0.86
23. Nitromethane	−28.6	101.2	35.87[f]	11.88	1.3812	46.3	1.137	−0.33	1.49	−0.99
24. Acetonitrile	−43.8	81.6	37.5[a]	11.48	1.3441	46.0	0.7857	−0.34	1.27	−1.65
25. 3-Pentanol	−75	115.3	13.02[e]	5.47	1.4103	45.7	0.8201	1.21	−0.33	−0.86
26. Dimethylsulfoxide	18.5	189.0	46.68	13.0	1.4783	45.0	1.101	−1.35	2.54	0.95
27. Aniline	−5.98	184.4	6.89[a]	5.04	1.4863	44.3	1.0217	0.90	−0.01	2.00
28. Sulfolane	28.5	287.3	43.3[f]	16.05	1.4920[f]	44.0	1.262[f]	–	3.29	2.21
29. Acetic anhydride	−73.1	140.0	20.7[g]	9.41	1.3904	43.9	1.0820	–	0.42	−0.87
30. 2-Methyl-2-propanol	25.8	82.4	12.47	5.54	1.3877	43.9	0.789	0.37	0.08	−0.37

31.	*N,N*-Dimethylformamide	−61	152.3	37.0	12.88	1.4269[d]	43.8	0.925[d]	−1.01	1.69	−0.50
32.	*N,N*-Dimethylacetamide	−20	166.1	37.78	12.41	1.4384	43.7	0.937[d]	−0.77	1.81	0.15
33.	Propionitrile	−92.8	97.4	27.2[a]	11.91	1.3658	43.7	0.782	0.16	0.68	−1.65
34.	1-Methyl-2-pyrrolidone	−24.4	204.0–204.8	−32.0	13.64	1.4700	42.2	1.026[d]	–	1.69	1.00
35.	Acetone	−94.7	56.3	20.70	9.54	1.3587	42.2	0.790	−0.24	0.17	−2.16
36.	Nitrobenzene	5.8	210.8	34.82	13.44	1.5500	42.0	1.204	1.85	1.03	2.47
37.	Benzonitrile	−12.8	191.1	25.20	13.51	1.5282	42.0	1.010	1.56	0.86	1.81
38.	1,2-Diaminoethane	11.3	117.3	12.9	6.34	1.4568	42.0	0.899	–	−0.07	0.60
39.	1,2-Dichloroethane	−35.7	83.5	10.36	6.20	1.4448	41.9	1.235	1.48	−0.65	−0.23
40.	2-Methyl-2-butanol	−8.8	102.0	5.82	5.7	1.4049	41.9	0.806	1.36	−0.45	−0.20
41.	2-Butanone	−86.7	79.6	18.51[a]	9.21	1.3788	41.3	0.835	0.29	−0.01	−1.57
42.	Acetophenone	19.6	202.0	17.39	9.87	1.5342	41.3	1.0281	1.58	0.51	2.24
43.	Dichloromethane	−95.1	39.8	8.93	5.17	1.4242	41.1	1.33	1.25	−1.23	−1.44
44.	1,1,3,3-Tetramethylurea	−1.2	175.2	23.45	11.58	1.4493[d]	41.0	0.969	–	1.12	0.88
45.	Hexamethylphosphoric triamide	7.2	235	29.6	18.48	1.4584	40.9	1.024[d]	0.28	2.26	1.61
46.	Cyclohexanone	−32.1	155.7	18.3[a]	10.04	1.4510	40.8	0.9478	0.81	0.41	0.42
47.	Pyridine	−41.6	115.3	12.4[h]	7.91	1.5102	40.2	0.982	0.65	−0.22	0.47
48.	Methyl acetate	−98.1	56.3	6.68	5.37	1.3614	40.0	0.933	0.18	−0.78	−2.10
49.	4-Methyl-2-pentanone	−84.0	116.5	13.11[a]	–	1.3957	39.4	0.7978	–	−0.97	−0.90
50.	1,1-Dichloroethane	−97.0	57.3	10.0	6.61	1.4164	39.4	1.176	1.79	−1.27	−1.20
51.	Quinoline	−14.9	237.1	9.00	7.27	1.6273	39.4	1.093	2.03	−0.29	3.13
52.	3-Pentanone	−38.9	1.2.0	17.00[a]	9.41	1.3923	39.3	0.8138	1.91	−0.03	−0.58
53.	Chloroform	−63.6	61.2	4.81[a]	3.84	1.4429[d]	39.1	1.48	1.92	−1.56	−0.59
54.	Triethylene glycol dimethyl ether	–	222	7.5	–	1.4233	38.9	–	–	0.12	0.97
55.	Diethylene glycol dimethyl ether	–	159.8	–	6.57	1.4097	38.6	–	–	−0.18	0.03
56.	Dimethoxyethane	−58	85	7.20	5.70	1.3796	38.2	0.8629	–	−1.11	−0.98
57.	1,2-Dichlorobenzene	−17.0	180.5	9.93	7.57	1.5515	38.1	1.305	3.38	−0.99	2.27
58.	Ethyl acetate	−84.0	77.1	6.02	6.27	1.3724	38.1	0.900	0.73	−0.44	−0.62
59.	Fluorobenzene	−42.2	84.7	5.42	4.90	1.4684[i]	38.1	1.023	2.27	−1.45	0.17
60.	Iodobenzene	−31.3	188.3	4.63[a]	4.64	1.6200	37.9	1.831	3.25	−1.46	2.74
61.	Chlorobenzene	−45.6	131.7	5.62	5.15	1.5248	37.5	1.106	2.84	−1.51	1.18
62.	Bromobenzene	−30.8	155.9	5.40	5.17	1.5571[d]	37.5	1.495	2.99	−1.40	1.81
63.	Tetrahydrofuran	−108.5	66	7.58	5.84	1.4072	37.4	0.889	0.46	−1.04	−1.57

Table 1. Continued.

Solvent (ordered after descriptor 6, $E_T(30)$)	Descriptors 1 m.p. (°C)	2 b.p. (°C)	3 ε_{25} °C	4 $\mu \cdot 10^{30}$ (C m)	5 n_D^{20}	6 $E_T(30)$ (kcal/mol)	7 $\rho \cdot 10^3$ (kg/m^3)	8 *log P*	*t*-values t_1	t_2
64. Anisole	−37.5	153.8	4.33	4.17	1.5170	37.2	0.996	2.11	−1.24	1.21
65. Phenotole	−29.5	170.0	4.22[a]	4.54	1.5074	36.4	0.967	2.51	−1.26	1.45
66. 1,1,1-Trichloroethane	−30.4	74.0	7.53[a]	5.24	1.4379	36.2	1.339	2.49	−1.53	0.03
67. 1,4-Dioxane	11.8	101.3	2.21	1.50	1.4224	36.0	1.034	−0.27	−0.67	−0.97
68. Trichloroethylene	−86.4	87.2	3.42[j]	2.7	1.4746[h]	35.9	1.464	2.29	−2.04	−0.28
69. Piperidine	−10.5	106.7	5.8[a]	3.97	1.4525	35.5	0.861	0.85	−0.92	0.28
70. Diphenyl ether	26.9	258.3	3.69[a]	3.87	1.4763[f]	35.3[f]	1.075	4.21	−1.39	3.79
71. Diethyl ether	−116.3	34.6	4.34[a]	4.34	1.3524	34.6	0.714	0.77	−1.67	−2.30
72. Benzene	5.5	80.1	2.28	0.0	1.5011	34.5	0.8787	2.15	−2.05	0.88
73. Diisopropyl ether	−85.5	68.3	3.88	4.20	1.3681	34.0	0.7251	2.03	−1.88	−1.29
74. Toluene	−95.0	110.6	2.38	1.43	1.4969	33.9	0.867	2.73	−2.46	0.24
75. Di-n-butyl ether	−95.2	142.2	3.08[a]	3.94	1.3992	33.4	0.7689	–	−1.75	−0.47
76. Triethylamine	−114.7	89.5	2.42	2.90	1.4014	33.3	0.7275	1.44	−1.95	−1.24
77. 1,3,5-Trimethylbenzene	−44.7	164.7	2.28[a]	0.0	1.4994	33.1	0.865	3.42	−2.40	1.36
78. Carbon disulfide	−111.6	46.2	2.64[a]	0.0	1.6280	32.6	1.263	–	−3.57	0.98
79. Carbon tetrachloride	−23.0	76.8	2.24[a]	0.0	1.4574[d]	32.5	1.59	2.83	−2.49	0.38
80. Tetrachloroethylene	−22.4	121.2	2.30	0.0	1.5057	31.9	1.623	2.60	−2.29	1.16
81. Cyclohexane	6.5	80.7	2.02[a]	0.0	1.4262	31.2	0.778	3.44	−2.61	0.57
82. *n*-Hexane	−95.3	67.8	1.88	0.0	1.3749	30.9	0.66	–	−3.00	−1.20

Values measured at other temperatures (°C), [a]20, [b]32, [c]28, [d]25, [e]22, [f]30, [g]19, [h]21, [i]15, [j]16.

studied phenomenon or reaction. In such situations, linear free energy relationships (LFERs) are likely to apply[2]. However, common to all attempts by this approach is that a rather large number of experimental observations have to be collected before an analysis can be made. This is hardly the case that exists in the initial screening phase of new synthetic methods in development. One way to proceed in this situation is to use a rather coarse-meshed screen to classify solvents according to their chemical and/or physical properties. This is often done intuitively, and chemists are accustomed to classify solvents according to: *chemical constitution* ("like dissolves like"), *Brønstedt acid – base properties* (protic, aprotic, acidic, basic), *Lewis acid – base properties* (EPD/EPA abilities, hard/soft acid-base properties), *physical properties* (polar, apolar), *chemical properties* (cation solvation, anion solvation, hydrogen bonding), *etc. etc.*

It is very common to refer to *solvent polarity* as a criterion for selection. However, this is not without complications: "*The characterization of a solvent by means of "polarity" is an unsolved problem since the term "polarity" has, until now, not been precisely defined*".[3] What is usually meant by the "polarity" of a solvent is its ability to interact by electrostatic forces with charged species in solution. Often the dipole moment, μ, the dielectric constant, ε, or some empirical solvent parameter is taken as a quantitative measure of "polarity". A selection based on one single property is clear-cut and simple: Solvents which span a large range in this property affords the desired selection. However, one is not *sure* that the property used as a probe is at all strongly related to the new reaction under study. Since various interactions are possible in solution, and the chemist does not know *a priori* which will be important, the same criticism applies to any other attempts to select solvents according to a single measured property, (linear correlation studies on various scales related to solvent "polarity" have been presented.[4] The only reasonable way to proceed in solvent selection is by using multivariate methods. An approach to solvent classification by factor analysis has been presented.[5] In this study, descriptors for 22 solvents were analyzed and the solvents were classified into ten different classes. The following descriptors were studied. E_T, *DN*, *AN*, dipole moment, molar refractivity, and *KIR*. However, the use of these descriptors will limit the number of solvents available to study. E_T is an empirical parameter which measures the transition energy at 25 °C of the long wave absorption band of standard pyridinium-*N*-phenoxide betain dye, see Ref.1 *DN*, donor number, which measures the nucleophilicity, is defined as the negative of the enthalpy (kcal/mol) of the interaction between the solvent and antimony pentachloride in dilute dichloromethane solution.[6] *AN*, acceptor number, measures the electrophilicity of the solvent and is obtained from the ^{31}P NMR chemical shift of triphenylphosphine oxide as measured in the solvent.[6] *KIR* is the Kirkwood function of the dielectric constant, $(\varepsilon-1)/(2\varepsilon+1)$. The selection must be made from a much larger number of possible candidate solvents in order to be of any practical value in a screening situation.

A study of weak interactions in the liquid state by means of principal components analysis has also been presented.[7] This is closely related to solute – solvent interactions when strong, specific interactions are absent.

DATA

There are about 30 different empirical solvent scales described[1,2,8] but few of these are available for a large number of solvents.[1]

In our study we have used the following descriptors: *1, melting point; 2, boiling point; 3, dielectric constant; 4, dipole moment; 5, refractive index; 6,* E_T*; 7, density* and, *8, log P*[9]. *Log*

P is the logarithm of the equilibrium distribution of the solvent between 1-ocotanol and water at 25 °C. These data are given in Table 1. Initially, we also intended to include solvent viscosity as a descriptor. Viscosity data are generally available. However, these data were obtained under conditions so different, that we considered them to be too inconsistent to be used in the data analysis described below. Data for descriptors *1–6* were taken from Ref. 1, Appendix 1A. Solvent densities, *7*, were compiled from standard handbooks,[10], *log P* values were compiled from the works of Hansch and co-workers.[11]

DATA ANALYSIS

Each set of solvent descriptors was scaled to unit variance. The SIMCA package [12] was used to fit a principal components (PC) model to the scaled data. This can be illustrated geometrically:

Assume that each descriptor defines a coordinate axis. The eight different axes will thus define an eight-dimensional space in which each solvent can be described by a point (with coordinates equal to the eight descriptors). The whole set of 82 different solvents will define a swarm of points in the eight-dimensional descriptor space.

Principal components analysis constitutes a projection of this swarm down to a space of lower dimensions in such a way that the first component vector (eigenvector) describes the direction through the swarm showing the largest variation in the data. The second eigenvector shows the second next largest variation, *etc.* The eigenvectors are mutually orthogonal. The principles are illustrated in three dimensions in Fig. 1.

The mathematical expression of a PC model will have a form:

$$x_{ik}=a_i+\sum_{j=1}^{A} b_{ij}t_{jk}+e_{ik} \qquad (1)$$

x_{ik} is the scaled value of the descriptor i for solvent k. The analysis corresponds to a least squares fitting of a straight line (A=1) or an A-dimensional hyperplane to the data points in the eight-dimensional descriptor space. The parameters a_i determine the center of the data set; the parameters b_{ij} are the direction coefficients (one for each variable and component)

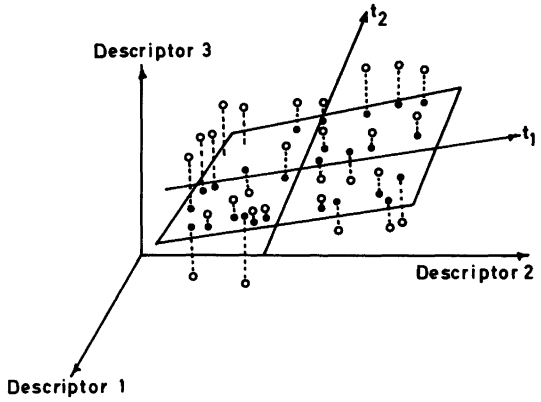

Fig. 1. Geometric illustration of the principles of PC modelling with three descriptors and two components.

of the line/hyperplane. For each solvent k, the parameters t_{jk} describe the position of the solvent point projected down to the model. Hence, t-values can be used to relate solvents to each other. The b-values (loadings) together with the residual variance, can give information of how much each descriptor variable contributes to the model.

An important result of this empirical modelling is that the *systematic variation* in the data can be described with fewer variables than in the original data set. Determination of the significant number of product terms (components) in eqn. 1 is made by cross validation. The principles for this have been given in detail elsewhere.[13]

RESULTS

A two components model of the whole data set in Table 1 is illustrated by the eigenvector projection in Fig. 2. The first component describes 29 % of variance in Table 1, and with the second component 51 % of the total variance is described. Introduction of a third component did not improve the explained variance and a three components model was not significant according to cross validation.[13] Different classes of solvents show groupings in the eigenvector plot and some examples of such groupings are shown in Fig. 3. The relative position of the solvent in Fig. 2 gives information on the similarities and dissimilarities between solvents. Similar solvents are close to each other in the eigenvector projection. A physical interpretation of the t-vectors in Fig. 2 is suggested from the b-plot in Fig. 4. The t_1 axis is mainly explained by the descriptors *3, 4* and 6 (dielectric constant, dipole moment, E_t) *i. e.* t_1 correlates well with the "polarity" concept discussed in the introduction. The t_2 axis is strongly correlated with descriptor *5* (refractive index). This means that polarizability is another important solvent property that shows a systematic variation and that "polarity" and polarizability are almost orthogonal and thus contribute independently of each other. Descriptor *7* (density) does not participate at all in the model and *1, 2, 8* (melting point, boiling point, *log P*) contribute to both components.

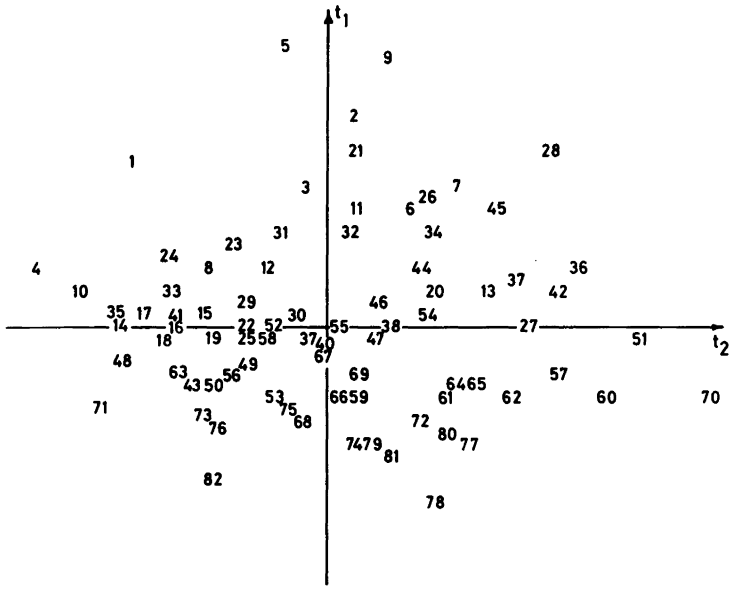

Fig. 2. Eigenvector projection of the solvent descriptor space.

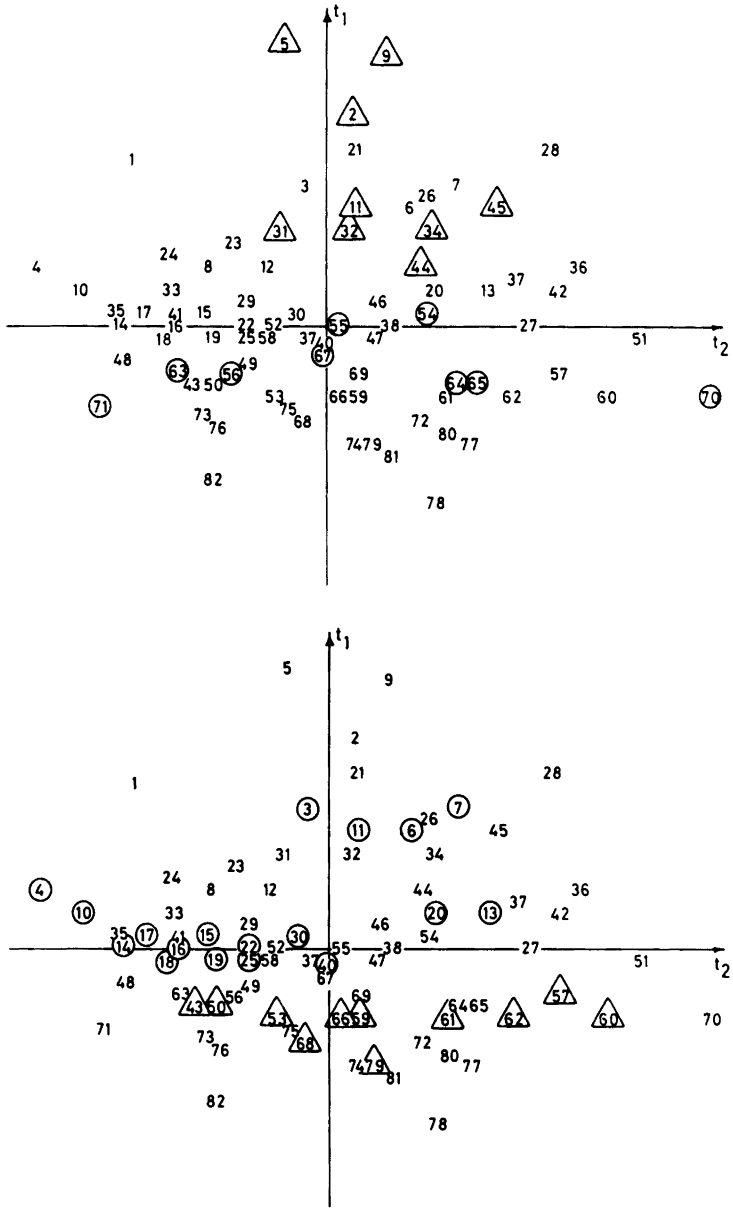

Fig. 3. Examples of groupings in the eigenvector projection. *3a*: Circles show monofunctional ethers, triangles show amides. *3b*: Circles show alcohols, triangles show halocarbons.

SOLVENT SELECTION

From the eigenvector projection in Fig. 2 we have a picture of the *systematic variation* in the "solvent space". When we have to make a choice in an experimental situation, we can use this information in different ways:

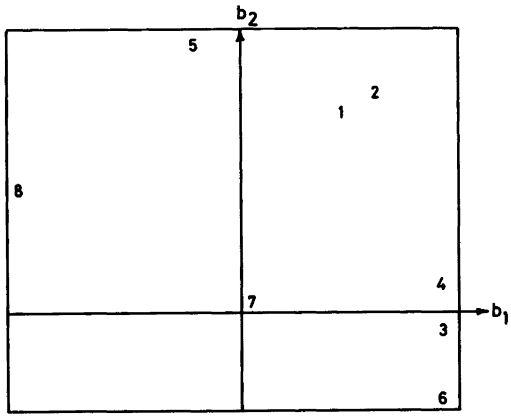

Fig. 4. A plot showing the contribution of each descriptor to the eigenvectors, *b*-plot. The loading b_2 is plotted versus b_1.

(a) *Select solvents that give a uniform coverage of the solvent space.* Such a selection can be obtained by taking solvents that form a regular lattice in the eigenvector projection. If a suggested solvent is incompatible with the reaction conditions take a nearby, more suitable, solvent.

(b) *Use a sequential simplex strategy* [14] *to explore the solvent space.* However, *t*-values are not continuous variables and it is not possible to "make reflections of the worst point" in a strict geometrical sense. Nevertheless, the principles can be applied in an approximate way: Select three solvents that form a simplex (triangle) in the eigenvector projection. As an example we can choose *N,N*-dimethylformamide, *31*; 2-butanol, *18,* and dioxane, *67.* Run the reaction and determine in which solvent the outcome is least favourable. The next solvent to be studied is chosen so that the remaining better solvents and the new one form a new simplex, oriented away from the discarded solvent in the first simplex *etc.* This is repeated until a suitable domain in the solvent space has been found. If, for instance, the reaction does not go well in dimethylformamide, *31,* try diisopropyl ether, *73,* or triethylamine, *76.* By the same principles 2-butanol, *18,* would be replaced by 1-methyl-2-pyrrolidone, *34,* and dioxane, *67,* by water, *1.* This strategy allows for a systematic search for better solvents for the reaction, by an interactive moving away from poor solvents, in the solvent space.

(c) *Select typical solvents from different solvent classes.* A common approach is to classify solvents as polar/apolar, protic/aprotic... *etc.* The eigenvector projection allows typical solvents from different classes to be selected, *i.e.* solvents that are not at the extreme ends of subgroupings in the solvent space, *cf.* Fig. 3.

(d) *Select solvents that are as dissimilar as possible to each other.* This will give a selection of solvents in the periphery of the domain and will ensure a large variation in solvent properties, see also (e) below.

(e) *Select solvents by an optimal design.* In the screening phase of new synthetic procedures no prior knowledge exists. Though we cannot explicitly say how solvent properties influence the reaction, we can assume that such an influence exists. The outcome of the experiment, *y*, is in some way related to solvent properties and we can assume a functional dependence:

$y=f$ (solvent properties)

The eigenvector projection shows the *systematic variation* of the solvent properties in Table 1 and the t-values *measures* this systematic variation. Hence, we can assume that:

$$y=f(t_1, t_2)$$

However, we do not know the nature of the function f but as an approximation we will attempt three different approximate models for f:

(*i*) f is described by a linear combination of solvent properties

$$y=c_0+c_1t_1+c_2t_2+\varepsilon \tag{2}$$

(*ii*) f is approximated by a synergistic model which includes a term to describe an interaction among solvent properties:

$$y=c_0+c_1t_1+c_2t_2+c_{12}t_1t_2+\varepsilon \tag{3}$$

(*iii*) f is approximated by a quadratic model in which the quadratic terms describe deviations from linearity:

$$y=c_0+c_1t_1+c_2t_2+c_{12}t_1t_2+c_{11}t_1^2+c_{22}t_2^2+\varepsilon \tag{4}$$

If we adopt a statistical experimental design approach, we can say that good test solvents are those from which an empirical model can be established with good precision. There are several different criteria for experimental design quality.[15] We will use a criterion of D-optimality,[16] which maximizes the overall precision of the estimated parameters. Different algorithms for the construction of D-optimal designs have been presented.[17] The design defines a set of solvents for which t_1 and t_2 are treated as independent variables. The tentative models can be written in matrix notation: $y=\mathbf{t}\,\mathbf{c}+\varepsilon$, where $\mathbf{t}=(1, t_1, t_2,\ldots)$ is a vector of variables in the model and $\mathbf{c}$ is a column vector of model parameters $\mathbf{c}'=(c_0, c_1, c_2\ldots)$. Let $\mathbf{T}$ be a matrix in which the rows define $\mathbf{t}$ vectors for the selected solvents. A design is D-optimal in which the information matrix determinant $|\mathbf{T}'\mathbf{T}|$ is as large as possible. Since $\sqrt{|\mathbf{T}'\mathbf{T}|}$ is inversely proportional to the joint confidence region for the estimated model parameters, a D-optimal design will maximize the overall precision. D-optimality is also independent of scaling of the variables. In our study we have used the algorithm by Fedorov.[17b] A computer program called NEMROD [18] was used for the calculation. To make a selection of solvents by this approach we have to specify the model and how many solvents are to the selected. The computer program then goes through all candidate solvents and iteratively searches those which give $|\mathbf{T}'\mathbf{T}|_{max}$. The result of this study is summarized in Table 2. It is seen in Fig. 2 that the selected solvents are found in the extreme periphery of the solvent space, *cf.* (d) above. In the case of a quadratic model one solvent near the center of the solvent space is also included. It is also seen that several test solvents are common to the selection for all three models, *i.e.* the selection of ten solvents of which nine are common to all three models. This indicates that even if a detailed knowledge is lacking of *how* different solvent properties influence the reaction, a rather small subset of test solvents can be used to study solvent "effects".

Table 2. Solvents selected by a D-optimal design.

Model	Number of test solvents	Selected solvents[a]
Linear	5	4, 5, 9, 70, 82
[Eqn. (2)]	6	4, 5, 9, 70, 71, 78
	7	4, 5, 9, 51, 70, 71, 82
	8	4, 5, 9, 51, 70, 71, 78, 82
	9	4, 5, 9, 28, 51, 70, 71, 78, 82
	10	1, 4, 5, 9, 28, 51, 70, 71, 78, 82
Synergistic	5	1, 5, 28, 70, 82
[Eqn. (3)]	6	1, 5, 28, 70, 71, 82
	7	1, 5, 28, 70, 71, 78, 82
	8	1, 5, 9, 28, 70, 71, 78, 82
	9	1, 4, 5, 9, 28, 60, 70, 71, 82
	10	1, 4, 5, 9, 28, 60, 70, 71, 78, 82
Quadratic	6	1, 9, 44, 70, 71, 78
[Eqn. (4)]	7	4, 5, 28, 46, 70, 78, 82
	8	1, 4, 5, 28, 46, 70, 71, 82
	9	1, 4, 5, 28, 46, 70, 71, 78, 82
	10	1, 4, 5, 9, 28, 46, 70, 71, 78, 82

[a] See *Table 1* for identification.

When a suitable solvent has been found by any of the principles discussed above, the next step will be to explore solvents located in the vicinity of the winning candidate in the eigenvector projection. This will allow it to meet other constraints (economy, ease of work-up procedures *etc.*) and may give directions towards optimization.

DISCUSSION

The results given above should not be regarded as a solution to the solvent selection problem, but merely as an approach to a solution. It is most likely that the picture of the solvent space given by PC eigenvector projection will change when other descriptors, not covered by this study, become generally available for a sufficiently large number of solvents. Another picture will emerge if the chemist has prior information of the reacting system *e.g.* the reaction involves strongly acid reagents or intermediates which are incompatible with basic solvents. Such information will reduce the number of candidate solvents to be considered and will produce PC models with less noise, *i.e.* more of the variance explained by the model. Anyhow, the strategies suggested will make it possible to systematically explore the solvent space and select test solvents which account for a variation in all properties considered. We can not give preference to any of the selection strategies (a–e) presented. It is, at present, a matter of judgement which one will be the appropriate in a given situation.

In this paper we have only discussed pure solvents. Of course there are innumerable ways to use solvent mixtures in organic synthesis. We suggest that pure solvents should be used in an initial screening of solvent "effects". Optimization of solvent mixtures can be handled later on in the study. Excellent reviews of useful strategies for optimization of mixtures by response surface modelling have been presented.[19]

CONCLUSIONS

Chemistry today tends to produce amounts of numbers and figures. Any single chemical system can be characterized by numerous measured properties and any single experiment can produce a multitude of numerical, measured, responses. The only reasonable way to handle this "information explosion" is to use computer-assisted multivariate strategies. The principles for the design of experiments in selecting the solvents, outlined in this paper, are an example of such strategies. These principles can be extended to cope with the general problem of selecting test objects in a screening situation. For instance, substrates, reagents, catalysts *etc.* can be characterized by a set of descriptors. Principal components analysis of these descriptors will reveal the systematic variation among test objects. A selection is then made on basis of this information. We can imagine yet another important application of this methodology to organic synthesis in the future *viz.* in the selection of test substrates to determine scope and limitations for new synthetic procedures.

CALCULATIONS

The calculations were carried out on a ZAMPO 8-bit CP/M microcomputer. Principal components modelling was obtained by the SIMCA program package[12] (SIMCA-3B version). SIMCA is an acronym Soft Independent Modelling of Class Analogy. The program is written in BASIC and versions for standard CP/M microcomputers are available from SEPANOVA, Östrandsvägen 14, S-122 43 Enskede, Sweden and from *Principal Data Components,* 2505 Shephgard Blvd, Columbia, Missouri 65201, USA. D-Optimal designs were constructed by a program called NEMROD[18] (New Efficient Methodology for Research using Optimal Designs). NEMROD is an interactive program for construction and analysis of experimental designs. It is written in FORTRAN.

Acknowledgements. We express our gratitude to LPRAI (*Laboratoire de Prospective Réactionnelle et d'Analyse de l'Information*), Aix-en-Provence, for the generous gift of the NEMROD program. Financial support from the Swedish Natural Research Science Council and from the National Swedish Board for Technical Development is gratefully acknowledged. We also thank Doc. M. Sjöström and Doc. S. Wold for stimulating discussions and Mr P. Power for linguistic assistance.

REFERENCES

1. a. Reichardt, C. *Solvent Effects in Organic Chemistry,* Verlag Chemie, Weinheim, New York 1979; b. Reichardt, C. and Harbusch-Görnert, E. *Justus Liebigs Ann. Chem.* (1983) 721.
2. Reichardt, C. *Angew. Chem. 91* (1979) 119.
3. Reichardt, C. *Justus Liebigs Ann. Chem.* (1983) 42.
4. Abboud, J.L., Kamlet, M.J. and Taft, R.W. *J. Am. Chem. Soc 99* (1977) 8325.
5. Chastrette, M. *Tetrahedron 35* (1979) 1448; Chastrette, M. and Carretto, J. *Tetrahedron 38* (1982) 1615.
6. Gutmann, V. *Electrochim. Acta 21* (1976) 661; Gutmann, V. and Wychera, E. *Inorg. Nucl. Chem. Lett. 2* (1966) 297; Mayer, V. and Gutmann, V. *Monatsh. Chem. 101* (1970) 912.
7. Cramer, R.D., III. *J. Am. Chem. Soc. 102* (1980) 1837, 1849.
8. Reichardt, C. In Ratajczak, H. and Orville-Thomas, W.J., Eds., *Molecular Interactions,* Wiley, New York 1982, Vol 3, pp. 241–282.

9. Hanisch, C. and Fujita, T. *J. Am. Chem. Soc. 86* (1964) 1616.
10. a. *Handbook of Chemistry and Physics,* Chemical Rubber Company, Cleveland 1972; b. Weissberger, A. *Technique of Organic Chemistry, Organic Solvents,* 2nd Ed., Interscience, New York, London 1955, Vol. 7.
11. Leo, A., Hansch, C. and Elkins, D. *Chem. Rev. 71* (1971) 525; Hansch, C. and Leo, A. *Pomona College Medicinal Chemistry Data Bank,* Pomona College, Claremont, CA 91711, USA.
12. a. Wold, S. *Pattern Recognition 8* (1976) 127; b. Sjöström, M. and Wold, S. In Kowalsky, B., Ed., *Chemometrics, Theory and Applications, Am. Chem. Soc. Symp. Ser. 52.,* American Chemican Society, Washington 1977; c. Albano, C., Dunn, W., III, Edlund, U., Johansson, E., Nordén, B., Sjöström, M. and Wold, S. *Anal. Chim. Acta Comput. Techn. Optim. 103* (1978) 429.
13. Wold, S. *Technometrics 20* (1978) 397.
14. Spendley, H.G.R., Hext, G.R. and Himsworth, F.R. *Technometrics 4* (1962) 441.
15. Nalimov, V.V. *9th European Meeting of Statisticians, Budapest 1972.*
16. Kiefer, J. *J. Roy. Statist. Soc. B 21* (1959) 273.
17. a. Mitchell, T.J. *Technometrics 16* (1974) 203, 211; b. Fedorov, V.V. *Theory of Optimal Experiments,* Academic, New York 1972.
18. Phan-Tan-Luu, R. and Mathieu, D. *NEMROD,* Laboratoire de Prospective Réactionnelle et d'Analyse de l'Information, Aix-en-Provence 1982.
19. a. Scheffe, H. *J. Roy, Statist. Soc. B 20* (1958) 344; b. Cornell, J.A. *Experiments with Mixtures,* Wiley, New York 1981; c. Snee, R.D. *Chemtech 9* (1979) 702.

Received April 13, 1984.