Principal Properties for Synthetic Screening: Ketones and Aldehydes

Rolf Carlson,[a,*] Michal P. Prochazka[a] and Torbjörn Lundstedt[b]

[a]Department of Organic Chemistry, University of Umeå, S-901 87 Umeå and [b]AB Leo, P.O. Box 839, S-201 80 Malmö, Sweden

Carlson, R. Prochazka, M. P. and Lundstedt, T., 1988. Principal Properties for Synthetic Screening: Ketones and Aldehydes. – Acta Chem. Scand., Ser. B 42: 145–156.

Commercially available ketones and aldehydes were characterized by readily accessible property descriptors. Principal components (PC) analysis of the data showed two significant components (principal properties). For the ketones, 88 % of the total variance was described by the PC-models, and for the aldehydes, 78 %. The cross-validation criterion was used to determine significant components.

Principles for the selection of test substrates in synthesis based on these models are briefly discussed.

To ensure maximum utility of new synthetic methods it is desirable that their scope and limitation can be established at an early stage. This implies that test substrates for investigating new methods should be selected to afford a sufficient spread in their properties. This is not a trivial problem when many properties are considered. We have recently described how analysis by multivariate statistical methods, viz. principal components (PC) analysis, of property descriptor data for a series of similar compounds can be used to reduce the complexity of the problem. PC-modelling allows the *systematic* variation in the properties over the whole series of compounds to be described by few (two or three) principal components, and the name "principal properties" has been suggested for these.[1] Principal properties have previously been reported for organic solvents,[2] Lewis acids,[3] amines in the Willgerodt reaction[4] and for amino acids.[1] A strategy for exploring the general scope of synthetic reactions has also been discussed.[5]

Organic synthesis is heavily dependent on the chemistry of the carbonyl group. Aldehydes and ketones are often relay targets in complex natural product syntheses. Hence, many new methods produced by synthesis are developed with simple carbonyl compounds as model substrates. In this paper we report a PC analysis of ketones and aldehydes in order to determine their principal properties, in the hope that our results will be useful to others in the field of organic synthesis. We are currently using these results in an extensive study on regioselectivity in the Fischer indole reaction, the results of which will be reported separately.

Data

Ketones. Eleven property descriptors were used to characterize the set of ketones. Data are summarized in Table 1. The set contains 79 items. One of them, 2(5*H*)-furanone (No. 43), is a lactone that was listed among ketones in a compilation of data found in the literature.[6] This compound should be an outlier in the PC modelling of ketones and was included as a test object for this purpose.

Aldehydes. Nine property descriptors were used to characterize a set of 113 aldehydes. Data are summarized in Table 2.

*To whom correspondence should be addressed.

Table 1. Property descriptors and PC scores of ketones.

Id.[a]	Ketone	Descriptors[b]											PC scores	
		1	2	3	4	5	6	7	8	9	10	11	t_1	t_2
1	1-Acetonapthone	170.21	11	1.628	1.12	302	6.0	12.9	151.97	8.23	–	1677.0	−3.66	−1.77
2	2-Acetonapthone	170.21	54	–	–	300.5	6.0	7.8	–	8.31	–	1674.6	−3.65	−0.97
3	Acetone	58.08	−94	1.3585	0.791	56	5.9	8.2	73.43	9.71	−203.8	1715.1	4.43	−1.45
4	Acetophenopne	120.15	19.5	1.5325	1.03	202	5.9	7.9	116.65	9.29	−195.7	1685.2	−1.09	−2.00
5	2-Adamantanone	150.22	257	–	–	–	5.7	9.5	–	–	–	1720.2	−3.67	4.15
6	1-Adamantyl methyl ketone	178.28	54	–	–	–	5.9	8.2	–	8.82	–	1701.4	−1.70	1.37
7	Anthrone	194.23	153	–	–	–	7.6	14.0	–	8.83	–	1658.6	−6.23	−0.96
8	Benzanthrone	230.27	169	–	–	–	6.1	12.8	–	–	–	1648.0	−7.93	−0.78
9	Benzophenone	182.22	50	–	–	305	6.0	7.0	–	9.14	–	1659.8	−4.53	−2.13
10	2-Butanone	72.11	−87	1.3788	0.805	80	5.8	7.3	89.58	9.54	−206.25	1717.5	3.77	−1.06
11	Butyrophenone	148.21	12	1.5195	1.021	221	5.9	8.2	145.16	–	–	1687.2	−1.52	−1.06
12	Cyclobutanone	70.0	–	1.4195	0.938	99	5.6	9.2	74.63	9.61	−207.9	1783.2	4.45	−0.95
13	Cyclobutyl phenyl ketone	160.22	–	1.547	1.05	–	6.0	14.4	152.59	–	–	1678.4	−2.35	−1.19
14	Cyclodecanone	154.25	24	1.482	0.958	–	5.9	6.9	161.01	–	−212.1	1701.6	−0.92	−0.04
15	Cyclododecanone	182.31	60	–	0.906	–	5.9	6.8	201.23	–	−208.4	1711.7	−1.57	1.72
16	Cycloheptanone	112.17	–	1.4611	0.951	179	5.9	6.9	117.95	9.14	−211.4	1702.1	0.53	−1.16
17	Cyclohexanone	98.15	−47	1.45	0.947	155	5.8	8.2	103.64	–	−208.5	1714.0	1.48	−1.39
18	Cyclohexyl phenyl ketone	188.27	56	–	–	–	6.0	14.3	–	–	–	1667.3	−3.86	−1.28
19	Cyclononanone	140.23	28	1.477	0.959	–	5.9	6.8	146.23	–	−214.2	1701.6	−0.54	−0.35
20	Cyclooctanone	126.2	40	–	0.958	196	5.9	6.8	131.73	9.00	−215.6	1700.6	−0.50	−0.76
21	Cyclopentadecanone	224.39	62	–	0.897	–	5.8	6.8	250.16	–	−208.6	1711.6	−2.77	3.10
22	Cyclopentanone	84.12	−51	1.4359	0.951	130.5	5.8	8.7	88.45	9.42	−213.6	1746.4	2.57	−1.17
23	Cyclopropyl methyl ketone	84.12	–	1.4241	0.849	114	5.9	7.2	99.08	9.46	–	1696.7	2.53	−1.42
24	1-Decalone	152.24	–	1.4917	0.989	–	5.9	6.9	153.93	–	–	1710.0	−0.73	−0.20
25	2-Decalone	152.24	–	1.49	0.979	–	5.9	6.9	155.51	–	–	1714.8	−0.58	−0.03
26	2-Decanone	156.27	4	1.4249	0.825	211	5.8	7.4	189.42	–	–	1718.9	0.07	1.81
27	3-Decanone	156.27	−4	1.4241	0.825	204.5	5.8	6.8	189.42	–	–	–	0.09	1.67
28	4-Decanone	156.27	–	1.4237	0.824	206.5	5.8	–	189.65	–	–	1714.0	−0.03	1.74
29	Decyl phenyl ketone	232.37	35	–	–	–	5.9	14.5	–	–	–	1688.9	−3.30	1.52
30	3,4-Dimethyl-acetophenone	148.21	–	1.538	1.001	243	6.0	7.9	148.07	–	–	1681.1	−1.93	−1.04

contd

Table 1. (contd)

Id.[a]	Ketone	Descriptors[b]											PC scores	
		1	2	3	4	5	6	7	8	9	10	11	t_1	t_2
31	2,4-Dimethyl-benzophenone	210.28	–	1.592	1.074	–	6.0	7.9	195.79	–	–	1622.4	−4.75	−0.31
32	2,5-Dimethyl-benzophenone	210.28	35	1.588	–	–	6.0	7.9	–	–	–	1665.1	−4.01	−0.98
33	3,4-Dimethyl-benzophenone	210.28	46	–	–	–	6.1	7.9	–	–	–	1656.4	−4.79	−1.73
34	2,6-Dimethyl-4-heptanone	142.24	–	1.4128	0.847	169	5.8	7.3	167.93	9.04	−215.5	1712.7	0.66	1.00
35	2,4-Dimethyl-3-pentanone	114.19	−80	1.3986	0.806	124	5.9	9.7	141.67	8.94	−213.1	1713.4	2.32	0.22
36	4,4-Dimethyl-2-pentanone	114.19	–	1.4037	0.809	127.5	5.9	7.4	141.15	9.23	−206.1	1716.8	1.96	0.40
37	2,2-Dimethyl-propiophenone	162.23	–	1.5084	0.97	220.5	6.0	10.4	167.25	–	–	–	−1.63	−0.14
38	Dodecyl phenyl ketone	260.42	46	–	–	–	5.9	14.5	–	–	–	1689.2	−4.13	2.39
39	3-Ethylacetophenone	148.21	–	1.5264	0.93	–	5.9	7.8	159.37	–	–	1685.4	−1.31	−0.42
40	4-Ethylacetophenone	148.21	−21	1.5293	0.993	–	5.9	7.9	149.25	–	−195.7	1682.7	−1.20	−1.13
41	4-Ethylcyclo-hexanone	126.2	–	1.4515	0.895	193	5.8	8.5	141.0	–	–	1717.6	0.57	0.07
42	Flavone	222.4	98	–	–	–	6.1	13.0	–	–	–	1646.9	−6.54	−1.74
43	2(5*H*)-Furanone	84.07	4.5	1.4692	1.185	–	5.6	5.7	70.95	–	–	–	–	–
44	9-Heptadecanone	254.46	52	–	–	–	5.9	6.8	–	–	–	1711.8	−2.85	4.24
45	2-Heptanone	114.19	−35	1.4085	0.82	149.5	5.9	8.6	139.26	9.18	–	1718.1	1.80	0.34
46	3-Heptanone	114.19	−39	1.4085	0.818	147.5	5.8	6.8	139.60	9.02	–	1715.4	1.80	0.29
47	4-Heptanone	114.19	−33	1.407	0.817	145	5.9	7.3	139.80	9.12	–	1713.3	1.75	0.27
48	Heptyl phenyl ketone	190.29	17	1.5077	–	–	6.0	14.6	–	–	–	–	−1.89	0.50
49	2-Hexanone	100.16	−57	1.4005	0.812	127	5.9	7.4	123.35	9.44	−206.8	1716.7	2.46	−0.15
50	3-Hexanone	100.16	–	1.4002	0.815	123	5.8	7.3	122.90	9.12	–	1714.9	2.34	−0.16
51	Hexanophenone	176.26	25.5	1.5105	0.958	265.2	5.9	14.5	183.99	–	–	1687.1	−2.26	0.38
52	Isobutyrophenone	148.21	–	1.5172	0.988	217	8.1	14.2	150.01	–	−202.8	1684.0	−1.48	−0.86
53	2-Methylbenzophenone	196.25	–	1.5958	1.083	–	6.1	8.0	181.21	–	–	1664.7	−4.41	−0.78
54	3-Methylbenzophenone	196.25	–	1.597	1.095	–	6.0	7.8	179.22	–	–	1660.8	−4.44	−0.95
55	4-Methylbenzophenone	196.25	57	–	–	326	6.1	7.8	–	–	–	1653.9	−5.26	−2.09
56	3-Methyl-2-butanone	86.13	−92	1.388	0.85	94.5	5.9	7.2	101.33	9.23	−209-3	1715.7	3.18	1.01
57	2-Methylcyclo-hexanone	112.17	–	1.4478	0.924	162.5	5.8	6.9	121.40	9.05	−210	1712.6	1.03	−0.75

contd

Table 1. (contd)

Id.[a]	Ketone	Descriptors[b]											PC scores	
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	t_1	t_2
58	4-Methylcyclo-hexanone	112.17	–	1.4142	0.914	170	5.9	9.0	122.72	9.16	−208.7	1717.5	1.10	0.55
59	5-Methyl-3-heptanone	128.22	–	1.4064	0.823	159.5	5.9	9.0	155.80	–	–	1714.1	1.17	0.74
60	5-Methyl-3-hexanone	114.19	–	1.4064	0.809	135	5.8	7.3	141.15	–	−211.8	1718.3	1.90	0.43
61	4-Methyl-2-pentanone	100.16	−80	1.396	0.8	117.5	5.9	7.2	125.20	9.42	−206	1717.0	2.76	−0.12
62	4-Methylpropiophenone	148.21	7	1.528	0.993	238.5	6.0	–	149.25	–	–	1685.6	−1.61	−0.90
63	10-Nonadecanone	282.51	56	–	–	–	5.8	6.8	–	–	–	1705.1	−3.94	4.43
64	2-Nonanone	142.24	−21	1.421	0.832	–	5.8	7.3	100.10	9.38	−206.4	1718.8	1.89	−0.33
65	3-Nonanone	142.24	–	1.4204	0.821	187.5	5.8	6.9	173.25	–	–	1716.8	0.57	1.32
66	5-Nonanone	142.24	−50	1.419	0.826	186.5	5.8	6.8	172.20	–	−208.3	1714.7	0.89	1.09
67	2-Octanone	128.22	−16	1.415	0.819	173	5.8	7.3	156.56	9.4	−208.2	1716.8	1.15	0.84
68	3-Octanone	128.22	–	1.415	0.822	167.5	5.8	6.9	155.99	9.1	–	1714.3	1.10	0.77
69	8-Pentadecanone	226.4	42	–	–	178	5.9	6.8	–	–	–	1715.7	−0.88	3.51
70	2-Pentanone	86.13	−78	1.3897	0.812	100.5	5.8	7.3	106.07	9.47	−208.8	1717.4	3.17	−0.65
71	3-Pentanone	86.13	−40	1.3924	0.814	102	5.8	8.9	105.81	9.22	–	1716.6	2.88	−0.57
72	1-Phenyl-2-butanone	148.21	–	1.5122	0.998	–	5.8	14.3	148.51	–	–	1713.0	−0.74	−0.42
73	2-Phenylcyclo-hexanone	174.24	55	–	–	–	5.8	14.4	–	–	–	1700.5	−1.68	1.20
74	4-Phenylcyclo-hexanone	174.24	79	–	–	–	5.9	13.2	–	–	–	1709.8	−1.59	2.22
75	Pinacolone	100.16	–	1.3964	0.801	106	5.9	8.8	125.04	8.88	–	1708.5	2.43	−0.20
76	2,2,6-Trimethyl-cyclohexanone	140.23	–	1.447	0.904	178.5	5.8	6.9	155.12	–	–	1707.1	0.17	0.22
77	2-Undecanone	170.3	11	1.428	0.825	231.5	5.8	7.4	206.42	–	–	1719.2	−0.45	2.31
78	3-Undecanone	170.3	12	1.4291	0.827	226	5.8	9.1	205.93	–	–	–	−0.52	2.16
79	6-Undecanone	170.3	15	1.428	0.831	228	5.8	6.8	204.93	–	–	1715.2	−0.53	2.17

[a]Identification number used in the score plots. [b]Descriptors: **1**, molecular mass (10^{-3} kg mol^{-1}); **2**, melting point (°C); **3**, refractive index; **4**, density (10^3 kg m^{-3}); **5**, boiling point (°C); **6**, **7**, wavelengths of maximum IR absorption (10^{-6} m); **8**, molar volume (10^{-6} m^3 mol^{-1}); **9**, ionization potential (eV); **10**, chemical shift of the carbonyl carbon in the ^{13}C NMR spectrum (ppm); **11**, carbonyl IR absorption (wavenumber). Descriptors **1–7** were taken from Ref. 6; **8** was calculated; **9** was taken from Ref. 12; **10** was taken from Ref. 13; **11** was taken from Ref. 14.

Table 2. Property descriptors and PC scores of aldehydes.

Id.[a]	Aldehyde	Descriptors [b]									PC scores	
		1	2	3	4	5	6	7	8	9	t_1	t_2
1	Acetaldehyde	44.05	−125	1.3316	0.788	21.5	5.8	7.4	1726.2	55.9	4.39	−0.92
2	4-Acetoxybenzaldehyde	164.16	–	1.5379	1.168	–	5.9	8.4	1763.1	140.55	−0.62	1.09
3	5-Acetoxymethyl-2-furaldehyde	168.5	54	–	–	–	5.8	6.0	1732.1	–	−1.91	1.45
72	Acrylaldehyde	56.06	−87	1.402	0.839	53	5.9	10.2	1696.2	66.82	2.84	−1.57
4	2-Amyl-3-phenylpropenal	202.25	−75	1.5571	0.97	288	6.0	6.1	–	208.51	−1.65	2.20
5	*o*-Anisaldhyde	136.15	38	–	1.127	238	5.9	6.2	1689.2	120.81	−1.27	−1.29
6	*m*-Anisaldehyde	136.15	–	1.5523	1.119	–	5.9	7.9	1702.7	121.67	−0.94	−0.71
7	*p*-Anisaldehyde	136.15	−1	1.5713	1.119	248	6.2	7.9	1683.3	121.67	−1.69	−1.03
8	9-Anthracenecarbaldehyde	206.24	104.5	–	–	–	6.0	13.7	1669.2	–	−5.78	−1.20
9	Benzaldehyde	106.12	−26	1.5454	1.044	181.5	5.8	8.3	1702.5	101.65	−0.19	−1.10
112	1,3-Benzenedicarbaldehyde	134.13	89	–	–	–	5.9	8.7	1695.6	–	−2.84	−1.51
73	3-Benzyloxybenzaldehyde	212.25	57	–	–	–	5.9	13.5	1694.8	–	−4.20	0.62
74	4-Benzyloxybenzaldehyde	212.25	73.5	–	–	–	7.9	13.6	1687.7	–	−4.75	0.13
75	3-Benzyloxy-4-methoxybenzaldehyde	242.47	62.5	–	–	–	5.9	7.9	1677	–	−5.62	0.42
76	4-Benzyloxy-3-methoxybenzaldehyde	242.27	64	–	–	–	7.9	8.8	1678.1	–	−5.61	0.46
77	2-Bromobenzaldehyde	185.03	21.5	1.596	1.585	230	5.9	13.3	1697.1	116.74	−3.20	−1.42
78	3-Bromobenzaldehyde	185.03	–	1.5935	1.587	229	5.9	8.4	1699.2	116.59	−3.25	−1.42
79	4-Bromobenzaldehyde	185.03	56.5	–	–	–	5.9	6.3	1696.6	–	−3.44	0.05
80	2-Bromo-3-phenylpropenal	211.06	67	–	–	–	5.9	9.0	1689.5	–	−4.53	0.25
81	5-Bromosalicylaldehyde	201.02	106.5	–	–	–	7.9	8.6	1674.6	–	−5.53	−1.07
82	5-Bromocrotonaldehyde	245.08	63.5	–	–	–	5.9	9.5	–	–	−5.73	0.45
83	6-Bromoveratraldehyde	245.08	150.5	–	–	–	6.0	7.9	1669.7	–	−7.70	−0.60
13	2-Butenal	70.09	–	1.4365	0.846	104	5.9	8.7	1691	82.85	1.85	−1.37
10	4-Butoxybenzaldehyde	178.23	–	1.53	1.031	285	5.9	6.2	1694.1	172.87	−2.09	0.67
11	Butyraldehyde	72.11	−96	1.379	0.817	75	5.8	–	1727.6	88.26	3.10	−0.30
84	4-Butylbenzaldehyde	236.36	–	1.4781	0.926	–	7.5	9.5	–	255.25	−1.89	3.56
85	3-(4-*t*-Butylphenoxy)benzaldehyde	254.33	–	1.5702	0.984	–	5.9	8.0	1701.5	258.46	−3.62	2.85
86	2-Chlorobenzaldehyde	140.57	11	1.5658	1.248	212	5.9	13.2	1697.9	112.64	−1.66	−1.18
87	3-Chlorobenzaldehyde	140.57	17.5	1.5645	1.241	213.5	5.8	8.4	1702	113.27	−1.65	−1.10
88	4-Chlorobenzaldehyde	140.57	45.5	–	–	213.5	5.9	12.2	1704.2	–	−1.80	−0.59
90	2-Chloro-5-nitrobenzaldehyde	185.57	76	–	–	–	6.5	7.4	1690	–	−2.47	−0.18
91	2-Chloro-6-nitrobenzaldehyde	185.57	70	–	–	–	6.6	7.4	1709.9	–	−3.29	−1.16
92	4-Chloro-3-nitrobenzaldehyde	185.57	64	–	–	–	5.9	7.4	1700.9	–	−3.33	0.62
93	5-Chloro-2-nitrobenzaldehyde	185.57	67	–	–	–	6.6	8.4	1697.3	–	−3.48	0.22
89	2-Chloro-3-phenylpropenal	166.61	32.5	–	–	–	5.9	8.9	1697.1	–	−3.65	0.01

contd

Table 2. (contd)

Id.[a]	Aldehyde	Descriptors [b]									PC scores	
		1	2	3	4	5	6	7	8	9	t_1	t_2
12	*trans*-Cinnamaldehyde	132.16	–	1.6221	1.048	248	6.0	8.9	1676.7	126.11	−1.98	−1.16
94	3-Cyanobenzaldehyde	131.13	76.5	–	–	210	5.9	14.8	1699.7	–	−2.22	−1.34
95	4-Cyanobenzaldehyde	131.13	99	–	–	–	5.9	8.3	1706.9	–	−2.62	−1.10
96	Decanal	156.27	–	1.428	0.83	208	5.8	6.8	1727.9	188.28	0.33	1.98
97	3,5-Dibromosalicylaldehyde	279.93	83	–	–	–	7.8	8.3	1682.2	–	−0.68	1.43
98	3,5-Di-*tert*-butyl-4-hydroxy-benzaldehyde	234.34	188	–	–	–	6.0	9.1	1667.9	–	−8.24	−1.26
99	2,3-Dichlorobenzaldehyde	175.01	65.5	–	–	–	5.9	8.0	1686.9	–	−3.66	−0.75
100	2,4-Dichlorobenzaldehyde	175.01	71	–	–	233	5.9	12.1	1695.4	–	−3.11	−0.54
101	2,6-Dichlorobenzaldehyde	175.01	70.5	–	–	–	5.8	12.9	1716.2	–	−2.88	0.67
102	3,4-Dichlorobenzaldehyde	175.01	42.5	–	–	247.5	5.9	8.4	1705.2	–	−2.54	0.31
103	3,5-Dichlorobenzaldehyde	175.01	65	–	–	–	8.5	11.6	1692.7	–	−3.48	−0.46
14	2,4-Dimethyl-2,6-heptadienal	166.27	–	1.4676	0.862	–	5.9	6.0	1689.7	192.88	−0.70	1.17
15	2,3-Dimethoxybenzaldehyde	166.18	50	–	–	–	5.9	7.9	1691.7	–	−2.98	−0.60
16	2,4-Dimethoxybenzaldehyde	166.18	70.5	–	–	–	6.2	7.9	1672.8	–	−3.96	−1.72
17	2,5-Dimethoxybenzaldehyde	166.18	50.5	–	–	–	7.8	8.4	1681.8	–	−3.28	−1.10
18	3,4-Dimethoxybenzaldehyde	166.18	43	–	–	281	6.3	7.9	1683.1	–	−3.33	−0.88
19	3,5-Dimethoxybenzaldehyde	166.18	46.5	–	–	–	6.3	8.6	1703.4	–	−2.55	0.01
20	2,4-Dimethoxy-3-methylbenzaldehyde	180.21	53	–	–	–	6.2	9.0	–	–	−4.45	−1.71
21	4,6-Dimethoxysalicyldehyde	182.18	69	–	–	–	6.1	8.6	1645.3	–	−5.16	−2.69
22	2,4-Dimethylbenzaldehyde	134.18	−9	1.5492	0.962	–	5.9	6.2	1694.5	139.48	−0.78	−0.33
23	2,5-Dimethylbenzaldehyde	134.18	–	1.5422	0.95	–	5.9	8.1	1693.6	141.24	−0.62	−0.26
24	2,2-Dimethyl-4-pentenal	112.17	–	1.4203	0.825	124.5	5.8	10.9	–	135.96	1.59	0.56
25	Diphenylacetaldehyde	196.25	–	1.5893	1.106	315	5.8	14.3	1724.1	177.44	−2.60	1.27
26	Dodecanal	184.32	–	1.4344	0.835	–	5.8	6.8	1728.2	220.74	−0.22	2.70
27	2-Ethoxybenzaldehyde	150.18	–	1.5422	1.074	–	5.9	8.1	1690.2	139.83	−1.28	−0.42
28	4-Ethoxybenzaldehyde	150.18	–	1.5584	1.08	255	6.3	8.0	1696.7	139.06	−1.63	−0.25
29	3-Ethoxy-4-methoxybenzaldehyde	180.21	52	–	–	–	7.9	8.8	1676.8	–	−3.82	−1.02
30	4-Ethoxy-3-methoxybenzaldehyde	180.21	59.5	–	–	–	7.9	8.8	1683.8	–	−4.54	−3.07
31	3-Ethoxysalicylaldehyde	166.18	67	–	–	264	8.0	13.6	1646.4	–	−4.54	−3.07
32	4-Ethylbenzaldehyde	134.18	–	1.538	0.979	221	5.9	12.1	1702.9	137.06	−0.75	−0.10
33	2-Ethylbutyraldehyde	100.16	–	1.4018	0.814	117	5.8	6.8	1730.9	123.05	2.13	0.51
113	2-Ethylhexanal	128.22	–	1.4155	0.822	–	3.4	5.8	–	155.99	1.33	1.13
34	Formaldehyde 37 % / Water	30.03	–	1.3765	1.083	–	3.6	9.7	–	27.73	2.91	−2.94
35	5-Formylsalicylaldehyde	150.13	108.5	–	–	–	5.9	8.6	–	–	−6.03	−5.28
36	2-Furaldehyde	96.09	−36	1.5243	1.16	162	6.0	13.3	1674.5	82.84	−0.25	−2.19

contd

Table 2. (contd)

Id.[a]	Aldehyde	Descriptors [b]									PC scores	
		1	2	3	4	5	6	7	8	9	t_1	t_2
37	3-Furaldehyde	96.09	–	1.493	1.111	–	5.9	8.7	–	86.5	0.28	−1.76
38	Heptanal	114.19	−43	1.4125	0.85	153	5.8	6.8	1727.4	134.34	1.32	0.63
39	*trans*-2-Heptenal	112.17	–	1.4473	0.857	–	5.9	10.2	1693.9	130.89	0.89	−0.22
40	*cis*-9-Hexadecenal	238.42	–	–	–	–	5.8	6.8	1711.1	–	−3.41	2.47
41	*cis*-11-Hexadecenal	238.42	–	1.4526	–	–	5.6	6.8	1728.5	–	−1.57	3.90
42	Hexanal	100.16	–	1.4035	0.834	131	3.7	5.8	1727	120.1	1.94	0.37
43	*trans*-2-Hexenal	98.15	–	1.4455	0.846	–	6.0	8.7	1693.1	116.02	1.26	−0.59
44	2-Hydroxy-4-methoxybenzaldehyde	152.15	42	–	–	–	6.2	8.2	1630.6	–	−4.29	−3.94
45	2-Hydroxy-3-methoxybenzaldehyde	152.15	4	1.5784	1.219	–	6.0	7.8	1630.6	124.82	−2.61	−2.13
46	3-Hydroxy-4-methoxybenzaldehyde	152.15	114	–	–	–	7.8	8.9	1673	–	−4.48	−2.41
47	5-Hydroxymethyl-2-furaldehyde	126.11	33.5	1.5627	–	–	6.0	9.7	1673.8	–	−1.71	−2.14
48	2-Hydroxy-1-napthalenecarbaldehyde	172.18	83.5	–	–	–	6.1	7.6	1645.7	–	−5.19	−3.04
49	2-Hydroxy-5-nitrobenzaldehyde	167.12	129	–	–	–	6.1	7.4	–	–	−7.14	−5.46
50	3-Hydroxy-4-nitrobenzaldehyde	167.12	130	–	–	–	6.3	8.0	1697.7	–	−4.45	−0.94
51	4-Hydroxy-3-nitrobenzaldehyde	167.12	141	–	–	–	5.9	7.9	1686.1	–	−6.01	−2.59
52	5-Hydroxy-2-nitrobenzaldehyde	167.12	167	–	–	–	7.5	7.6	1670.8	–	−6.01	−2.59
53	Isobutyraldehyde	72.11	−65	1.3725	0.794	63	5.8	–	1738	90.82	3.11	−0.13
54	4-Isopropylbenzaldehyde	148.21	–	1.5298	0.977	235.5	6.0	12.0	1702.6	151.69	−1.01	0.30
55	Isovaleraldehyde	86.13	–	1.3882	0.803	90	5.8	6.8	1727.6	107.26	2.61	0.09
56	Mesitaldehyde	148.21	14	1.5522	1.005	237	5.9	6.2	1687.8	147.47	−1.60	−0.25
104	2-Methylpropenal	70.05	−81	1.416	0.847	69	5.9	–	1695.7	82.75	2.42	−1.23
57	7-Methoxy-3,7-dimethyloctanal	186.3	–	1.4374	0.877	–	5.8	9.2	–	212.43	−0.29	2.55
58	2-Methyl-3-phenylpropenal	146.19	–	1.6045	1.047	–	6.0	14.4	1682.3	139.62	−1.79	−0.73
59	2-Methylundecanal	184.32	–	1.4321	0.83	171	5.8	6.8	1728.9	222.07	0.05	2.66
60	2-Methylvaleraldehyde	100.16	–	1.4067	0.808	119.5	5.8	6.9	1726.3	123.96	2.02	0.43
106	1-Naphthaldehyde	156.18	1.5	1.652	1.15	–	5.9	13.0	1689.3	135.81	−2.37	−0.82
107	2-Naphthaldehyde	156.18	60.5	–	–	–	5.9	13.3	1695	–	−2.83	−0.77
111	*trans*-2-Nonenal	140.23	–	1.4531	0.846	–	5.9	10.3	1694.2	165.76	0.16	0.63
105	Nonanal	142.24	–	1.424	0.827	–	5.8	6.8	1727.6	171.99	0.92	1.52
110	Octanal	128.22	14	1.4183	0.821	171	5.8	6.8	1727.6	156.18	0.55	0.99
61	*trans*-2-Pentenal	84.12	–	1.4414	0.86	80.5	5.9	8.7	–	97.81	1.99	−0.70
62	Phenylacetaldehyde	120.15	−10	1.5293	1.027	195	5.8	14.3	1723.7	116.66	−0.24	−0.28
63	(±)-3-Phenylbutyraldehyde	148.21	–	1.5179	0.997	–	5.8	14.2	–	148.66	−0.66	0.20
64	2,3-Diphenylpropenal	208.26	45	–	–	–	6.0	8.8	1666.7	–	−4.70	−0.79
65	D,L-2-Phenylpropionaldehyde	134.18	–	1.5176	1.011	–	5.8	14.2	1724.2	132.72	−0.04	0.17
108	Piperonal	150.13	36	–	–	264	5.9	7.9	1688.1	–	−2.69	−0.93

contd

Table 2. (contd)

Id.[a]	Aldehyde	Descriptors [b]									PC scores	
		1	2	3	4	5	6	7	8	9	t_1	t_2
66	Propionaldehyde	58.08	−81	1.365	0.805	48	5.8	7.2	1733.3	72.15	3.53	−0.61
67	Salicylaldehyde	122.12	1.5	1.5719	1.146	197	6.0	7.6	1664.9	106.56	−1.48	−1.91
109	Tetradecanal	212.38	24	–	–	–	5.8	6.8	1728.9	–	−2.51	2.61
68	Tridecanal	198.35	–	1.484	0.835	–	5.8	6.8	1728.4	237.54	−0.61	3.10
69	10-Undecenal	168.28	–	1.4427	0.81	–	5.8	11.0	1727.4	207.75	0.12	2.34
70	Undecanal	170.3	–	1.4322	0.825	–	5.8	6.8	1728.3	206.42	0.15	2.35
71	Valeraldehyde	86.13	−92	1.3942	0.81	103	3.7	5.8	1726.2	106.33	2.59	0.09

[a]Identification number used in the score plot. [b]Descriptors: **1**, molecular mass (10^{-3} kg mol^{-1}); **2**, melting point (°C); **3**, refractive index; **4**, density (10^3 kg m^{-3}); **5**, boiling point (°C); **6**, **7**, wavelengths of maximum IR absorption (10^{-6} m); **8**, carbonyl IR absorption (wavenumber); **9**, molar volume (10^{-6} m^3 mol^{-1}). For references to the descriptors, see footnote to Table 1.

Methods

The principal properties were determined by PC projection of the data in Tables 1 and 2. The SIMCA program package was used for calculations.[7] An advantage of the SIMCA package is that it can handle incomplete data matrices. Detailed accounts of the principles of PC modelling have been given previously.[7,8]

Prior to the analysis each descriptor variable was scaled to unit variance over the whole data set, so that different units of measurement would not distort the variance and bias the result. The number of significant components was determined by cross validation.[9]

PC Modelling and results

Ketones. The consistency of the data was checked as follows: In a first run, all variables were included. Three sub-sets of ketones were selected from Table 1, allowing, respectively, a maximum of one, two, and three missing data for each ketone. These sub-sets were used as training sets in the PC analysis. Two components were extracted, which accounted for ca. 60 % of the total variance. PC projections are shown in Fig. 1. The cross-validation criterion and the calculated modelling power for each variable indicated that descriptor **10** (^{13}C NMR shift for the C=0 group) did not show a systematic variation over the data set. The same was also found for descriptors **6** and **7** (strongest IR absorption bands). Recalculation of the PC model after deletion of these variables, and with training sets allowing, respectively, one, two and three missing data for each ketone, afforded significant two-component models which accounted for 80–82 % of the total variance. When the remaining ketones were projected to the model, some outliers were found (ketones Nos. 5–9, 12, 22, 42–44, 63 and 69). These were not included in the training set for the final modelling. Descriptor **10** was also deleted. The reason for this was that data were missing for 47 of the 78 ketones and that cross validation indicated only a small contribution from this variable to the model. A significant two-component model was obtained which described 88 % of the total variance. Ketones Nos. 5, 7, 12, 22 and the lactone No. 43 were projected outside the 95 % confidence region of the model when all the objects in Table 2 were fitted to the model. The PC

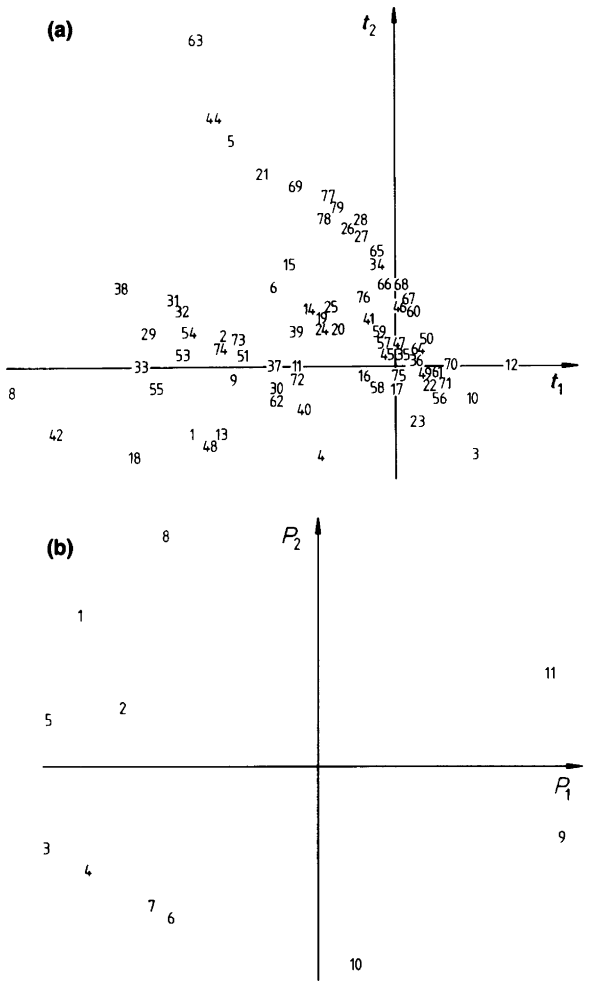

Fig. 1. PC projections for ketones based on 11 descriptor variables. (a) Score plot, (b) loading plot.

projections are shown in Fig. 2. The PC scores (t values) which measure the principal properties are listed in Table 1.

Aldehydes. In a preliminary analysis, 53 simple aldehydes (not containing multiple functionalities) were selected from Table 1. PC decomposition showed two significant components (88 % of variance accounted for). As found for the ketones, descriptors **6** and **7** did not contribute to the model and were deleted.

In the next step, training sets were selected from Table 2 allowing, respectively, a maximum of one and two missing data for each item. PC analysis yielded almost identical results for these training sets, viz. 78 and 77 % of the total variance accounted for.

The principal properties were obtained by projecting the whole data set in Table 2 onto the model obtained from the training set with a maximum of two missing data. The following aldehydes were outside the 95 % confidence limits of the model: Nos. 2, 3, 50–52, 95, 98 and 112.

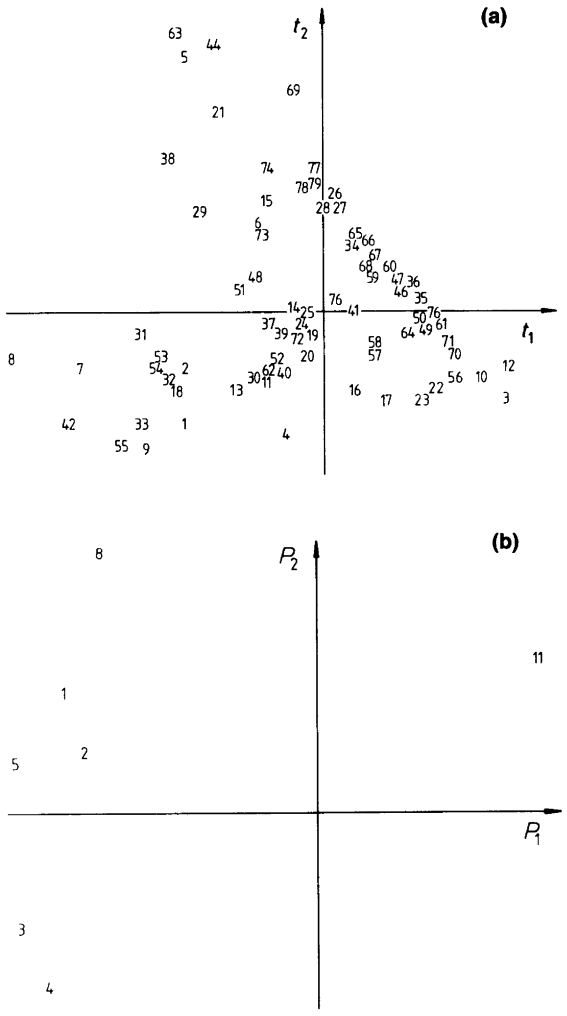

Fig. 2. PC projections for ketones based on seven descriptors. (a) Score plot, (b) loading plot.

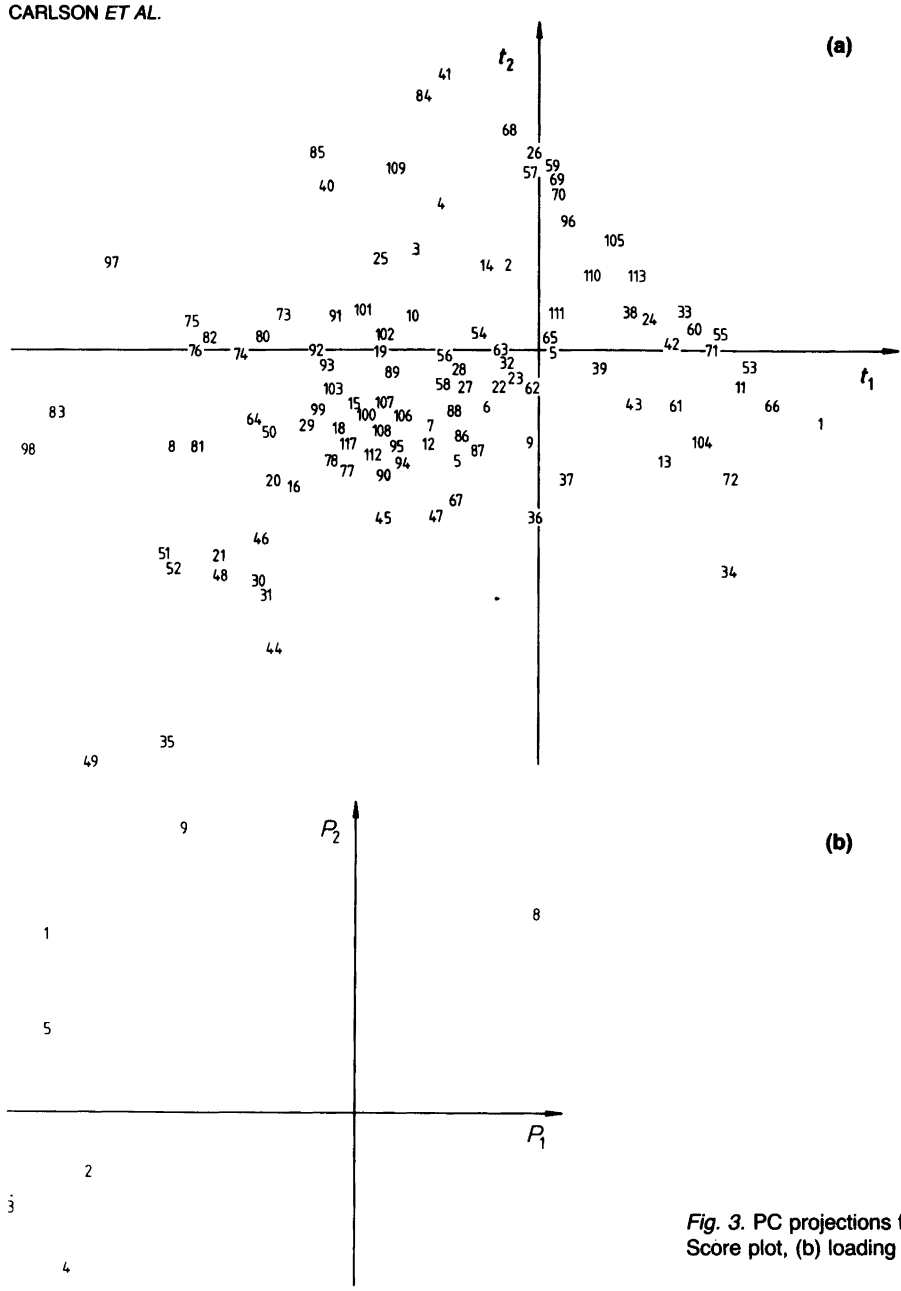

Fig. 3. PC projections for aldehydes. (a) Score plot, (b) loading plot.

These outliers are excluded from PC projections shown in Fig. 3. The PC scores are given in Table 2.

Discussion

Why bother about principal properties? Quantum chemical modelling can adequately describe isolated small molecules and account for very simple reactions in the gas phase. With larger molecules, certain features can be described but the complexity increases rapidly with the number of atoms involved and more or less drastic assumptions must be made to make calculations pos-

sible. It is, however, well beyond the scope of present-day theoretical chemistry to give any useful description of even simple synthetic reactions in which the reactants are dissolved in a solvent. This means that a purely theoretical approach is excluded when analyzing synthetic methods. This must be done on the basis of experimental observations. A common approach to the selection of test systems is to base it on an assumed reaction mechanism. With new reactions, however, mechanistic details are unknown; any selection of test substrates based on a reaction hypothesis that later proved to be false would then have involved wasted effort. We therefore suggest that selection of test candidates should be made on the basis of the principal properties. The underlying principles can be explained as follows:

A number of factors act on a chemical compound to determine its behaviour in a chemical reaction. Certain factors are related to the experimental conditions (temperature, concentration of reactants etc.), and information on such factors can only be gained from properly designed experiments in which such factors are varied.[10] Other factors refer to the reaction system (substrate, reagent, solvent) and involve, for example, effects of electronic distribution (field effects, delocalization), intermolecular forces (e.g. in relation to aggregation), solute – solvent interactions, steric hindrance to attack of a reagent, and conformational effects. Common to these factors is that they are dependent on intrinsic properties of the molecular species involved and interaction between such properties. Unfortunately, such intrinsic properties cannot be measured directly. What can be observed are macroscopic *manifestations* of these intrinsic properties.[11]

When a number of properties are measured for a series of similar compounds, it is often found that several property descriptors are correlated to each other, i.e. they are related to the same intrinsic property. In PC analysis, correlated descriptors will contribute in a similar way to the PC vectors. Totally independent properties will be described by different PC vectors. The contribution of each of the descriptors to the direction of the PC vectors is given by the loadings (see Figs. 1b, 2b and 3b). Since PC vectors are mutually orthogonal, we can assume that they portray *independent* and *different* intrinsic properties. We can therefore use the PC scores (t values) as measures of the intrinsic properties. We can also use the t values for comparison of compounds.

Selection of test compounds. The PC score projections (Figs. 2a and 3a) show the systematic variation in the principal properties. We can use this information and select test compounds using two principles:

(A) Select test objects that are projected far from each other and on the periphery of the projection. This will give a test set in which the variation in the principal properties is maximized.

(B) Select test objects that are uniformly spread over the projection. This will give a test set in which the variations in principal properties are evenly spread over the test objects.

The bulk properties used as descriptors convey information on intermolecular forces. Such forces will also play a role in solute – solvent interactions. The aldehydes classified as outliers in the PC model are all aromatic aldehydes which are either phenolic or are substituted with strong electron-withdrawing groups. With these aldehydes the bulk properties are strongly influenced by hydrogen bonding and strong polar interactions. This indicates that weak van der Waals interactions are responsible for a systematic variation for the remainder of the aldehydes. Ability to participate in van der Waals interaction is therefore considered as an important intrinsic property.

The position of the carbonyl absorption in the IR spectrum depends on the electronegativity of the substituents at the carbonyl group. In ketones, a strain factor also intervenes and this is the reason why cyclobutanone (No. 12) and cyclopentanone (No. 22) were classified as outliers.

However, there are important factors that are not covered by the descriptors used in this study. Such factors are, for example, the presence of bulky groups or branching of the carbon skeleton close to the carbonyl group, and the presence of other functionalities that might possibly interfere with the desired reaction. It is therefore suggested that selection by principal properties according to (A) or (B) above should be used as a "coarse-meshed screen" to select a sub-set of po-

tential test candidates with a sufficient spread in properties. A final selection is then made from this sub-set taking other factors into account.

Data. When the chemist is faced with the problem of selecting test substrates, it is likely that he/she resorts to commercial sources for supply of chemicals rather than to synthesizing a series of model compounds. In this study we have largely used data on commercially available chemicals. We believe that this will be sufficient in most cases.

Two problems are encountered whenever data are compiled for a large series of compounds: (1) Literature data are not consistent. Scientific papers rarely describe more than a handful of compounds; to obtain data for series of compounds, data must be taken from different sources. However, the methods of measurement used by one group are seldom identical to the methods used by other groups. (2) Data are not available for all compounds, i.e. some descriptors are missing. This is not a major problem in the data-analytic method we have used since the SIMCA program tolerates missing data if they are scattered over the whole data set.

Conclusions

Elaborating a reaction into a synthetic method will usually entail determination of the scope of the reaction with regard to substrate variation. This necessitates selection of test substrates to cover a range of variation in properties that may interfere with the reaction. In this context, principal properties are useful tools since they permit all factors to be considered prior to any actual experimental run. Another advantage is that selection based on principal properties is less likely to be afflicted by personal bias.

Calculations

The calculations were carried out on IBM PC/XT, Toshiba M 1500 or Toshiba M 1100 (16-bit) micro-computers. The SIMCA program package (SIMCA 3-B version) was used to establish the PC models. The program is available from *SEPANOVA AB,* Östrandsvägen 14, S-112 43 Enskede, Sweden, or from Principal Data Components, Shepard Blvd., Columbia, Missouri 65201, USA.

Acknowledgements. Financial support to this study from the Swedish Natural Research Science Council and from the National Swedish Board for Technical Development is gratefully acknowledged. We also thank Mr. Stephen Swanson for linguistic assistance.

References

1. Hellberg, S., Sjöström, M., Skagerberg, B. and Wold, S. *J. Med. Chem. 30* (1987) 1126.
2. Carlson, R., Lundstedt, T. and Albano, C. *Acta Chem. Scand., Ser. B 39* (1985) 79.
3. Carlson, R., Lundstedt, T., Nordahl, Å. and Prochazka, M. *Acta Chem. Scand., Ser. B 40* (1986) 522.
4. Lundstedt, T., Carlson, R. and Shabana, R. *Acta Chem. Scand., Ser. B 41* (1987) 157.
5. Carlson, R. and Lundstedt, T. *Acta Chem. Scand., Ser. B 41* (1987) 164.
6. *The Aldrich Microfiche Library of Chemical Indices,* Aldrich Chemical Company Inc., 1987.
7. (a) Wold, S. *Pattern Recognition 8* (1976) 127; (b) Sjöström, M. and Wold, S. In: Kowalski, B. R., Ed., *Chemometrics, Theory and Applications,* American Chemical Society Symposium Series 52, American Chemical Society, Washington 1977.
8. (a) Joliffe, I. T. *Principal Components Analysis,* Springer, New York 1986; (b) Wold, S., Albano, C., Dunn, W. J., III, Edlund, U., Esbensen, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W. and Sjöström, M. In: Kowalski, B. R., Ed., *Chemometrics, Mathematics and Statitics in Chemistry,* D. Riedel Publ. Co., Dordrecht, Holland 1984.
9. Wold, S. *Technometrics 20* (1978) 397.
10. Box, G. E. P., Hunter, W. S. and Hunter, J. S. *Statistics for Experimenters,* Wiley, New York 1978.
11. Wold, S., Albano, C., Dunn, W. J., III, Edlund, U., Eliasson, B., Johansson, E., Norden, B. and Sjöström, M. In: Jöreskog, K. G. and Wold, H., Eds., *Systems Under Indirect Observation. Causality – Structure – Prediction, Part II,* North-Holland Publ. Co., Amsterdam 1982, pp. 178–190.
12. Levin, R. D. and Lias, S. G. *Ionization Potentials and Appearance Potential Measurements (1971–1981),* National Standard Reference Data Series 71, National Bureau of Standards, USA 1982.
13. Breitmaier, E. and Voelter, W. *^{13}C NMR Spectroscopy: Methods and Application,* Verlag Chemie, Weinheim 1974.
14. Pouchert, C. J. *The Aldrich Library of FT-IR Spectra,* Aldrich Chemical Company Inc., 1985.

Received September 14, 1987.