

Pattern Recognition Search for Basic Regularities in the Stability of Complex Hydrides. Part 3. Monosubstituted Complexes ABH_3D

SVANTE WOLD^a and OLDRICH STROUF^b

^a Research Group for Chemometrics, Institute of Chemistry, Umeå University, S-901 87 Umeå, Sweden and ^bInstitute of Inorganic Chemistry, Czechoslovak Academy of Sciences, 25068 REZ near Prague, Czechoslovakia

Complex hydrides ABH_3D (A=alkali atom, B=group III A atom and D an organic or inorganic ligand; e.g. $LiAlH_3Me$ and KBH_3F) were described by means of 1 variable for A, 2 for B and 12 for D. Alternatively 8, 8 and 12 variables, respectively, were used in a comparative analysis.

SIMCA pattern recognition analyses of two classes of complexes, stable ($N=26$) and unstable ($N=9$) result in a good classification (33 correct of 35, 28 of these 33 uniquely classified). The stabilities of another 18 complexes are predicted.

In part 1 of this series¹ we tried to find relations between the stability and structure of complex hydrides ABH_nD_{4-n} . Here A is an alkali metal (Li, Na, K, Rb or Cs), B is a group III A atom (B, Al, Ga, In or Tl) and D a substituent other than hydride, for instance alkyl or halide. The methodology was based on describing each of the structural fragments A, B and D by means of a number of theoretical or semi-empirical variables, such as valence radii and ionization potentials of the elements A and B, and electronic and size descriptors of D. Thereafter the resulting multivariate data set was analyzed by means of SIMCA pattern recognition to find regularity "patterns", in the data sets of the classes "stable" and "unstable" complexes.

In part 2² we discussed the possible causes of the limited success in the first trial. We proposed that a partitioning of the analysis into separate subanalyses of complexes ABH_4 , ABH_3D , ABH_2D_2 and $ABHD_3$, respectively, might give better and more easily interpreted results.

In part 2 complexes of ABH_4 were analyzed. It was shown that the 8 variables used to describe A and the 8 variables used for B could be confined to only 3 variables with little loss of information.

In the present article we analyze the subgroup ABH_3D using 12 variables to describe the substituent D. For the atoms A and B we use both the full description with 8+8 variables used in parts 1 and 2 and the contracted description with the 1+2 variables obtained from the part 2 analysis.

METHODOLOGY

The philosophy of pattern recognition as well as the SIMCA method have been discussed in detail in parts 1 and 2. Separate principal components (PC) models were derived for the two classes (I=stable and II=unstable) from the training set data. The possibility of an asymmetric data structure (see part 2) was considered but not found. The dimensionality of the two PC models was estimated using cross-validation. Irrelevant variables (low modelling power) were deleted and a final analysis was made using only the relevant variables. The complexes in the test set were fitted to the two final PC models and assigned to the class corresponding to the best-fitting model. The classification rate of the training set was checked by deleting parts of the training set in the same way as in part 2.

DATA

A thorough literature search provided stability data for 35 complexes of the formula ABH_3D . The

Table 1. Complexes, substituent number (Table 2), assignment and RSD-values for fit to class model I and II respectively. An asterisk on the assignment indicates a unique classification.

No.	Class	Name	D	Assignment	SD _I	SD _{II}
3	I	LiBH ₃ Me	2	I*	0.38	0.66
5	I	LiBH ₃ NMe ₂	64	I*	0.20	0.84
6	I	LiBH ₃ OMe	42	I, II	0.38	0.35
7	I	LiBH ₃ OBu- <i>t</i>	49	I*	0.42	0.74
8	I	LiBH ₃ OCHPh ₂	51	I*	0.46	1.9
9	I	LiBH ₃ PH ₂	151	I*	0.34	0.64
24	I	LiBH ₃ Et	3	I	0.43	0.66
25	I	NaBH ₃ Pr- <i>n</i>	4	I	0.56	0.74
26	I	NaBH ₃ O-Me	42	I, II	0.38	0.35
27	I	NaBH ₃ OC(Me)=NPh	57 ^a	I*	0.57	1.4
28	I	NaBH ₃ CN	121	I*	0.19	1.4
29	I	NaBH ₃ CONHMe	92	I*	0.20	1.3
30	I	NaBH ₃ PH ₂	151	I*	0.34	0.64
47	I	NaBH ₃ NMe ₂	64	I*	0.20	0.84
49	I	KBH ₃ NMe ₂	64	I*	0.20	0.84
50	I	KBH ₃ CN	121	I*	0.19	1.4
51	I	KBH ₃ CONHMe	92	I*	0.20	1.3
52	I	KBH ₃ CONMe ₂	93	I*	0.26	1.5
54	I	KBH ₃ PH ₂	151	I*	0.34	0.64
55	I	KBH ₃ F	101	I*	0.32	0.52
68	I	LiAlH ₃ OCHPh ₂	51	I*	0.61	1.6
87	I	NaAlH ₃ Bu- <i>iso</i>	8	I*	0.42	0.74
88	I	NaAlH ₃ C≡CC ₄ H ₉ - <i>n</i>	31	I*	0.60	1.3
89	I	NaAlH ₃ OBu- <i>t</i>	49	II, (I)	0.70	0.56
123	I	NaGaH ₃ C ₈ H ₁₇ - <i>n</i>	14	I*	0.50	1.5
127	I	KGaH ₃ C ₈ H ₁₇ - <i>n</i>	14	I*	0.50	1.5
132	II	LiBH ₃ OCH ₂ CH ₂ OMe	52	II*	1.2	0.24
133	II	LiBH ₃ SH	111	I, II	0.42	0.29
143	II	LiAlH ₃ OMe	42	II*	0.79	0.38
144	II	LiAlH ₃ OEt	43	II*	0.88	0.36
145	II	LiAlH ₃ OPr- <i>iso</i>	45	II*	0.73	0.41
146	II	LiAlH ₃ OBu- <i>t</i>	49	II, (I)	0.70	0.56
155	II	NaAlH ₃ Et	3	I, II	0.67	0.60
156	II	NaAlH ₃ OCH ₂ CH ₂ NMe ₂	53	II*	1.2	0.28
157	II	NaAlH ₃ OCH ₂ CH ₂ OMe	52	II*	1.3	0.41
53		KBH ₃ COOEt	83	I*	0.43	1.3
166		LiBH ₃ CN	121	I*	0.17	1.4
167		LiBH ₃ NC ₄ H ₄	36 ^a	I*	0.42	1.5
168		LiBH ₃ OH	41	I, (II)	0.30	0.54
178		NaBH ₃ COOH	81	I*	0.19	1.1
179		NaBH ₃ COOMe	82	I*	0.22	1.1
180		NaBH ₃ OH	41	I, (II)	0.30	0.54
181		NaBH ₃ OPh	55 ^a	II, (I)	0.60	0.44
184		NaBH ₃ F	101	I, (II)	0.29	0.52
199		KBH ₃ NEt ₂	66	I*	0.28	1.2
200		KBH ₃ OH	41	I, (II)	0.30	0.54
209		RbBH ₃ N(Bu- <i>n</i>) ₂	66 ^a	I*	0.28	1.2
211		LiAlH ₃ Me	2	II, (I)	0.77	0.55
212		LiAlH ₃ Et	3	II, (I)	0.67	0.49
213		LiAlH ₃ (C ₂ H ₃)	21	I, (II)	0.68	0.78
216		LiAlH ₃ Cl	102	II*	0.97	0.69
253		NaAlH ₃ OMe	42	II*	0.79	0.31
279		LiGaH ₃ Me	2	II*	0.83	0.60

^a Approximate values derived from a similar substituent with the number given.

detailed references are available as supplementary material, Data base 2, from S.W. Of these, 26 were characterized as stable and 9 as unstable (see parts 1 and 2 for discussion of the stability assignment). These complexes constitute the training set. In addition, 28 ABH_3D complexes from the literature with little or no stability information available were included as a test set. Table 1 shows the formulas of the included complexes.

For each complex, the atoms A and B were described in two alternative ways. In the first, the same 8+8 variables were used as in parts 1 and 2. In the second, one variable for A and two for B were used, the variables obtained from the analysis described in part 2. These latter 3 variables were shown in part 2 to contain 95% of the information of the original 8+8 variables describing A and B.

The substituent D was described by the set of 12 variables shown in Table 2. This set contains variables related to the size of D (Taft's E_s , Verloop's L , B_1 , B_2 , B_3 and B_4), its hydrophobic character (Hansch's π), electronic properties (MR = Molecular refractivity, Hammett's σ , *meta* and *para*) and two indicators describing electron donor/acceptor properties of atoms 1 and 3 to 4 in the substituent chain. These 12 variables have been found relevant in recent studies of relations between chemical structure and derived properties (biological activities).^{3,4}

Taft's E_s is a substituent descriptor derived from the substituent's influence on the rate of acidic and alkaline ester hydrolysis.⁶ E_s measures the steric influence of the substituent but also contains a polar, electronic part.^{7,8}

Verloop's parameters are derived from quantum mechanical calculations.⁹ L is a measure of the length of the substituent and B_1 to B_4 four different measures of its width.

Hansch's π is a measure of the hydrophobic (lipophilic) character of the substituent.^{10,11} π is derived from the distribution of model compounds between water and a hydrophobic phase, usually octanol.

The molecular refractivity, MR, is supposed to measure the polarizability of the substituent.^{10,11} Hammett's σ parameters measure the inductive effect of a substituent as revealed in aromatic reaction series.^{12,13,14}

The donor-acceptor indicators of atom 1 and 3-4 are simply 1 for donor atoms (O, N and halogens) -1 for acceptors (CN, carbonyl groups) and 0 otherwise (see Table 2).

In conclusion, the structures of the complexes were described by 28 and 15 variables in the two alternatives.

DATA ANALYSES AND RESULTS

The two alternative data sets were analyzed separately in exactly the same way except that the 28 variable results were not validated. Only the results of the 15 variable set are tabulated in detail since the two analyses give very similar results.

a. Scaling. To give each variable the same weight, the data were first normalized to zero mean and unit variance over the training set (auto-scaling).¹⁵ The scaling parameters are given in Table 3.

b. Cross validation. According to cross-validation,⁵ two principal components were needed in both class models and both analyses. When variables with low modelling power were deleted, four principal components in class 1 and two principal components in class 2 were found to be significant in the 15 variable case.

c. Deletion of irrelevant variables. In the 28 variable case, the following variables of the substituent D had both low modelling and discrimination power and were deleted: σ , *para* and *meta*, E_s , Verloop's L , B_1 and B_2 (D-variables 2,3,5,7,8 and 9 respectively).

In the 15 variable case, the following variables were deleted on the same grounds: No. 1 (describing atom A), Nos. 11 and 12 (Verloop's B_1 and B_2 of the substituent D).

d. Principal components (PC) analyses. Separate PC models were fitted to the two training set class data matrices. As found by cross-validation, 2-component models were used for both classes in the 28 variable case (6 variables deleted in the final analysis). A 4-component model was used for class 1 and a 2-component model for class 2 in the 15 variable case (3 variables deleted). In the latter analysis the residual standard deviations (RSD) for the two classes were both 0.41 compared to the SD of 1.0 in the original data. Thus, the models describe about 60% of the SD or 85% of the variance of the data.

The resulting PC parameters for the 15 variable case are shown in Table 3. From the β -values of class 1 and 2, it is seen that 8 of the variables are grouped two and two together, namely 2 and 3 (variables describing B), 5 and 6 (D; σ , *meta* and *para*), 7 and 10 (D; MR and Verloop's L) and 13

Table 2. Values used to describe the substituent D; main sources Ref.'s 9 and 11.

No.	Substituent name	PI	σ_m	σ_p	MR	E_s	D/A	V.L	V.B ₁	V.B ₂	V.B ₃	V.B ₄	D/A
2	Me	0.56	-.06	-.14	5.65	-1.24	0.00	3.00	1.52	1.90	1.90	2.04	0.00
3	Et	1.02	-.08	-.13	10.30	-1.31	0.00	4.11	1.52	1.90	1.90	2.97	0.00
4	Pr-n	1.55	-.07	-.10	14.96	-1.60	0.00	5.05	1.52	1.90	1.90	3.49	0.00
5	Pr-iso	1.53	-.07	-.15	14.96	-1.71	0.00	4.11	2.04	2.76	3.16	3.16	0.00
6	Bu-n	2.13	-.07	-.12	19.59	-1.63	0.00	6.17	1.52	1.90	1.90	4.42	0.00
7	Bu-sec	2.04	-.07	-.15	19.59	-2.37	0.00	3.05	1.90	2.76	3.16	3.49	0.00
8	Bu-iso	2.10	-.07	-.12	19.59	-2.17	0.00	5.05	1.52	1.90	3.16	4.21	0.00
9	Bu-tert	1.98	-.09	-.15	19.62	-2.78	0.00	4.11	2.59	2.86	2.86	2.97	0.00
10	Pent-n	2.64	-.08	-.15	24.24	-1.64	0.00	7.11	1.52	1.90	1.90	4.94	0.00
11	CHMeCHMe ₂	2.64	-.07	-.12	24.19	-2.37	0.00	5.05	2.11	2.76	3.16	3.99	0.00
13	Oct-n-straight	4.00	-.08	-.15	38.10	-1.57	0.00	10.27	1.52	1.90	1.90	7.33	0.00
14	Oct-n-compact	4.00	-.08	-.15	38.10	-2.03	0.00	5.82	1.52	2.17	4.75	6.85	0.00
15	CH ₂ CH ₂ OCH ₃	0.31	-.07	-.13	16.68	-2.01	0.00	6.03	1.52	1.90	1.90	3.81	1.00
16	CHMeCH ₂ CH ₂ OCH ₃	0.50	-.07	-.15	21.30	-2.50	0.00	6.99	1.90	2.76	3.16	3.81	1.00
17	CH ₂ CH ₂ OCH ₂ CH ₃	0.70	-.07	-.13	21.30	-1.63	0.00	6.99	1.52	1.90	1.90	4.74	1.00
21	Vinyl	0.82	.05	-.02	10.99	-1.40	-1.00	4.29	1.60	1.60	2.00	3.09	0.00
22	Cyclopentyl	2.14	-.05	-.10	22.02	-1.75	0.00	4.97	2.04	2.86	2.86	3.98	0.00
23	Cyclohexyl	2.51	-.08	-.15	26.69	-2.03	0.00	6.17	2.04	3.16	3.16	3.49	0.00
24	Cycloheptyl	2.90	-.08	-.15	31.34	-3.27	0.00	7.00	2.04	3.50	3.50	4.20	0.00
27	Phenyl	1.96	.04	.05	25.36	-3.82	-1.00	6.28	1.70	1.70	3.11	3.11	0.00
28	PhMe ₃ -2,4,6	3.16	.04	.05	39.30	-3.82	0.00	7.28	3.00	3.00	4.00	4.00	0.00
30	CH ₂ Ph	2.01	-.08	-.09	30.01	-1.62	1.00	3.63	1.52	3.11	3.11	6.02	-0.30
31	C≡C-Bu-n	2.40	.15	.15	24.04	-1.60	-1.00	8.70	1.70	1.70	3.16	4.21	0.00
36	N ₂ C ₃ H ₃	-0.25	.15	.15	24.04	-3.82	-1.00	5.97	1.65	1.77	3.13	3.16	0.00
41	OH	-0.67	.02	-.22	2.85	-0.55	1.00	2.74	1.35	1.35	1.35	1.93	0.00
42	OMe	-0.02	.10	-.12	7.87	-0.55	1.00	3.98	1.35	1.90	1.90	2.87	0.00
43	OEt	-0.38	.10	-.12	12.47	-0.86	1.00	4.92	1.35	1.90	1.90	3.36	0.00
45	OPr-iso	0.36	.10	-.12	17.06	-1.02	1.00	4.59	1.35	1.90	3.16	3.61	0.00
49	OBu-tert	1.00	.10	-.12	22.00	-2.09	1.00	4.92	1.35	2.86	2.86	3.94	0.00
51	OCHPh ₂	3.62	.10	-.12	50.00	-2.09	1.00	8.20	2.04	3.11	3.11	3.11	-1.00
52	OCH ₂ CH ₂ OMe	-0.03	.10	-.12	18.68	-1.80	1.00	6.84	1.35	1.90	1.90	4.71	1.00
53	OCH ₂ CH ₂ NMe ₂	0.03	.10	-.12	23.50	-2.00	1.00	6.84	1.35	1.90	3.25	4.71	1.00
54	OCHO	-1.00	.26	.20	7.50	-1.00	1.00	3.93	1.35	1.60	1.60	3.68	0.00
55	OCOMe	-0.64	.26	.16	12.47	-1.50	1.00	4.87	1.35	1.90	1.90	3.68	0.00
57	OCOPh	1.46	.21	.13	32.30	-2.00	1.00	8.15	1.70	1.70	1.84	4.40	-1.00
64	NMe ₂	0.18	-.10	-.32	15.55	-1.71	1.00	3.53	1.50	2.56	2.80	2.80	0.00
66	NEt ₂	1.18	-.12	-.35	24.85	-2.78	1.00	4.96	1.50	2.56	3.42	3.42	0.00

and 14 (D ; Verloop's B_3 and B_4). The remaining variables, 4, 8, 9 and 15 (D ; π , E_s , and donor/acceptor property of atom 1 and 3–4, respectively) show independent behaviour over the two classes.

e. Classification of the training set. The fit of each complex in the training set to the two final class models of the 15 variable case (3 variables deleted) is shown in Table 1. The 28 variable analysis gives a very similar picture. 6 complexes (Nos. 6, 26, 89, 133, 146, 155) are not uniquely classified; they are close to both class models. Three of these (Nos. 6, 26, 89) are slightly closer to the "wrong" model.

f. Validation. The classification rate found in the previous section was checked as described in part 2. The training set was divided into 4 groups. Four analyses were made, each time making one of the 4 groups a "pretended" test set with known expected class assignment. In each analysis a 4-component model was estimated from the reduced class 1-data set and a 2-component model from the reduced class 2-data set. The "pretended" test sets were then classified using these models.

The results show that the rate found in the previous section is reliable. Of those complexes belonging to class 1, three were ambiguously classified, two of them closer to the "correct" class model (Nos. 26 and 55) and one of them closer to the class 2 model (No. 6). No. 89 was uniquely assigned to the wrong class. Of the complexes in class 2, two were ambiguously but "correctly" classified (Nos. 146 and 155) and one uniquely classified to the wrong class (No. 133).

In conclusion, a worst expected rate of 5 ambiguous and 2 erroneous classifications out of 35 is found, a most encouraging reliability.

g. Classification of the test set. The final models of the 15 variable case (3 variables deleted) give the classification of the test set shown in Table 1. Of the 18 complexes, 10 are uniquely classified and 8 with more or less ambiguity.

h. Graphical representation. An eigenvector projection¹⁶ of the whole data set (15 variables, none deleted) is shown in Fig. 1. Class 1 (stable) is seen to be more spread out than class 2 (unstable). The figure shows that the two classes are fairly well



Fig. 1. Eigenvector projection (first two eigenvalues) of the 15 variable data set. Filled symbols denote complexes in the training set, open symbols complexes in the test set. Squares indicate stable and circles unstable complexes.

separated already in a two-dimensional projection.

i. Comparison of the 28 variable and 15 variable analyses. The two analyses give almost exactly the same results with respect to the classification of the complexes. In the interpretation of the results, however, they differ in that all 8 variables relating to atom A are retained as relevant in the 28 variable analysis (*a*-28) but the single variable describing atom A is deemed irrelevant in the 15 variable analysis (*a*-15). The reason why the *A*-variables are retained in *a*-28 is probably the same as why they are relevant in the analysis of ABH₄ in part 2. They show a substantial correlation pattern when atom A is varied over Li, Na and K in class 1 training set. This is seen from the parameters β of *a*-28; the main part of the second component describes this variation. When this component is "lifted out" and represented as a single variable in *a*-15, this variable becomes irrelevant because it shows no correlation with the variables describing *B* and *D*. Hence *a*-15 gives results which are easier to interpret in a straight - forward manner; the conclusion that the atom A has little influence on the stability is obtained also from *a*-28 on closer inspection.

PREDICTION FOR UNMADE COMPLEXES

By means of the parameters in Table 3, predictions can be made for yet untried complexes. This is done as follows, illustrated using NaGaH₃CN as an example:

1. Look up the descriptor values for atom B in Table 3 in part 2. Ga has the values -3.0 and 1.8.

2. Look up the descriptor values for D in Table 2 in the present paper. CN has the values shown in Table 4. In case D is not included in Table 2, the pertinent values must be looked up in more extensive tabulations where the Pomona College Data bank,¹¹ Verloop's original article⁹ and a recent compilation by Exner¹³ are recommended sources.

3. Form the 12-dimensional vector from the B and D values and scale it by subtracting \bar{y}_i and dividing by S_i in Table 3 in the present article. This gives the second row in Table 4.

4. Fit the two class models to the scaled data vector using the regression

$$y_i^* - \alpha_i = \sum_{a=1}^A \beta_{ia} t_a + \varepsilon_i$$

Here y_i^* denotes the scaled data-value of variable *i*. The parameters α_i and β_{ia} for $a=1$ to *A* are found in Table 3 for the two classes. Due to the orthogonality properties of the vectors β_a , the t_a -values are determined simply by

$$t_a = \sum_{i=1}^{12} (y_i^* - \alpha_i) \beta_{ia}$$

The residuals ε_i are then calculated as

$$\varepsilon_i = y_i^* - \alpha_i - \sum_{a=1}^A \beta_{ia} t_a$$

The residual SD for each class is calculated from ε_i as

Table 4. Predictions for NaGaH₃CN.

<i>i</i>	1	2	3	4	5	6	7	8
<i>y</i>	—	-3.0	1.8	-0.57	0.62	0.71	6.33	-0.51
y^*	—	-1.82175	-0.92200	-0.90239	2.75401	2.70656	-0.99915	1.50843
ε_1	—	-1.365	-0.533	-0.800	0.360	-0.066	-0.019	-0.506
ε_2	—	-1.416	-0.416	-0.694	2.665	2.781	-0.107	0.281
θ (I)	2.846	-3.101	-0.554	2.806				
θ (II)	2.614	-0.687						
<i>i</i>	9	10	11	12	13	14	15	SD
<i>y</i>	-1.0	4.23	—	—	1.60	1.60	0	—
y^*	-1.73351	-0.41713	—	—	-1.12748	1.41019	0.0	—
ε_1	0.722	-0.315	—	—	-0.279	-0.577	-0.379	0.739
ε_2	-2.339	0.640	—	—	-0.237	-0.591	0.780	1.56

```

100 N=12
110 DIM Y(N), A(N), B(4,N), T(4)
120 PRINT "DATA INPUT AND SCALING, GIVE Y, MEAN, SD"
130 FOR I=1 TO N
140 PRINT I
145 INPUT Y(I), A(I), B(1,I)           Input  $y_i, \bar{y}_i, s_i$ 
150 Y(I)=(Y(I)-A(I))/B(1,I)           Scaling  $(y_i - \bar{y}_i)/s_i$ 
160 PRINT "SCALED Y=", Y(I)
170 NEXT I
180 PRINT "GIVE NO OF COMPS, A"
190 INPUT M                             Input no of components
200 PRINT "GIVE ALFA AND BETAS"
210 FOR I=1 TO N                         Input  $\alpha_i, \beta_{ia}$  of one class.
220 PRINT "I="; I
230 INPUT A(I)
240 FOR J=1 TO M
250 INPUT B(J,I)
260 NEXT J
270 NEXT I
280 FOR J=1 TO M
290 T(J)=0
300 FOR I=1 TO N                         Regression-coefficients
310 T(J)=T(J)+(Y(I)-A(I))*B(J,I)          $t_a = \sum_i (y_i - \alpha_i) * \beta_{ia}$ 
320 NEXT I
330 PRINT "TETA(“;J;”)=", T(J)
340 NEXT J                               Print  $t_a$ 
350 S=0
360 PRINT "RESIDUALS"
370 FOR I=1 TO N                         Calculate and print
380 D=Y(I)-A(I)                          residuals
390 FOR J=1 TO M
400 D=D-B(J,I)*T(J)
410 NEXT J                               Calculate sum of squares
420 S=S+D↑2
430 PRINT I,D
440 NEXT I
450 D=SQR(S/(N-M))
460 PRINT "SS=";S;"SD=";D               Print sum of squares
470 STOP                                  and RSD
480 END

```

Fig. 2. BASIC program scaling and fitting a data vector y to the class model specified by the given input parameters. For the second class, restart with RUN 180 (the y -values are the same). For a second data-vector to be run with the same class-parameters, restart with RUN 120 and break at line 190 (at the request of A). Then restart with RUN 280.

$$S^2 = \sum_{i=1}^{12} e_i^2 / (12-A)$$

Table 4 lists the residuals ε_i for both class models and NaGaH₃CN together with the t_a -values and residual SD. It is seen that the residual SD is much smaller for class 1 than for class 2 (0.74 and 1.56, respectively). Hence NaGaH₃CN is predicted to be stable. The simple calculations listed above are readily performed on a programmable calculator

or a minicomputer using, for instance, the BASIC-program listed in Fig. 2.

DISCUSSION

The classification results can, together with the picture of Fig. 1, be interpreted as follows: Three main factors seem to influence the stability of ABH₃D complexes; in order of importance:

1. The donor/acceptor ability of atom 1 in the D substituent; donors being destabilizing (e.g. OH, OMe, NMe₂) and acceptors being stabilizing (e.g. -COOMe, CN). Alkyl-substituents seem to be intermediate. This is clearly seen in Fig. 1 where all complexes at the lower right, far from class 2, have D = acceptors.

2. The size of the D-substituent, in particular, its width. Large size seems to be stabilizing. Complexes with large D-groups lie to the lower left in Fig. 1.

3. The position of atom B in the periodic system; B being preferable to Al. This factor might be important only for small D-groups; No. 123, NaGaH₃-oct(n) is stable while No. 279, LiGaH₃Me is classified as unstable.

It is interesting to note that the analysis deems the nature of the A-atom as irrelevant to the stability. Note, however, the difference in stability between Nos. 89 and 146. We also note that donor-atoms at position 3 or 4 in D seem to have no stabilizing effect. In view of the stabilizing effect of acceptors in position 1, it might be interesting to try acceptor-groups out in the D-chain, say CN or carbonyl groups.

Finally we note the much improved classification rate obtained in both part 2 and the present case as compared with part 1. In part 1 a classification rate of about 75% was obtained compared with validated rates of 100% in part 2 and 94% (33/35) in the present case including ambiguous but correct assignments. We note that the division into subclasses gives no loss of information in the present case since the chemist knows if a complex is ABH₄ or ABH₃D — he would not be interested in determining the structure of the complex.

Acknowledgement. A grant from the Swedish Natural Science Research Council has given valuable support.

REFERENCES

1. Strouf, O. and Wold, S. *Acta Chem. Scand. A* 31 (1977) 391.
2. Wold, S. and Strouf, O. *Acta Chem. Scand. A* 33 (1979) 463.
3. Dunn, W., Wold, S. and Martin, Y. *J. Med. Chem.* 21 (1978) 922.
4. Dunn, W. and Wold, S. *J. Med. Chem.* 21 (1978) 1001.
5. Wold, S. *Technometrics* 20 (1978) 397.
6. Taft, R. W. In Newman, N. S., Ed., *Steric Effects in Organic Chemistry*, Wiley, New York 1956, Chapter 13.
7. Shorter, J. In Chapman, N. B. and Shorter, J., Eds., *Advances in Linear Free Energy Relationships*, Plenum, London 1972, Chapter 2.
8. Sjöström, M. and Wold, S. *J. Chem. Soc. Perkin Trans. 2*. In press.
9. Verloop, A., Hoogenstraaten, W. and Tipker, J. In Ariens, E. J., Ed., *Drug Design*, Academic, New York 1977, Vol. V.
10. Hansch, C., Leo, A., Unger, S. H., Kim, K. H., Nikiatani, D. and Lien, E. J. *J. Med. Chem.* 16 (1973) 1207.
11. Pomona College Data Bank, c/o Leo, A., Pomona College, Claremont, California 91711, U.S.A.
12. Hammett, L. P. *Physical Organic Chemistry*, 2nd Ed., McGraw-Hill, New York 1970.
13. Exner, O. In Chapman, N. B. and Shorter, J., Eds., *Correlation Analysis in Chemistry*, Plenum, London 1978, Chapter 10.
14. Wold, S. and Sjöström, M. In Chapman, N. B. and Shorter, J., Eds., *Correlation Analysis in Chemistry*, Plenum, London 1978, Chapter 1.
15. Harper, A. M., Duerwer, D. C., Kowalski, B. R. and Fashing, J. L. In Kowalski, B. R., Ed., *Chemometrics, Theory and Application*, ACS Symposium Series 52, 1977.
16. Kowalski, B. R. and Bender, C. F. *J. Am. Chem. Soc.* 95 (1973) 686.

Received February 19, 1979.