

Pattern Recognition Search for Basic Regularities in the Stability of Complex Hydrides. Part 2. Unsubstituted Complexes ABH_4

SVANTE WOLD^a and OLDRICH STROUF^b

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

Hydrides ABH_4 (A=alkali atom and B=Group III A atom; *e.g.* $LiAlH_4$) were described by means of 8 variables for A and 8 variables for B. An asymmetric SIMCA pattern recognition analysis of the resulting 16×15 matrix for 15 complexes known to be stable classifies these 15 stable complexes and 2 unstable complexes correctly. The stability of 8 yet unmade hydrides is predicted to be low.

In our previous paper¹ (part 1) we tried to find regularities in the stability-instability of complex hydrides with the formula ABH_nD_{4-n} where A is an alkali metal, B a metal in Group III A and D a substituent other than hydride, for instance halide or alkyl. Complexes in a learning set of 95 stable and unstable hydrides were described by means of structural data for A, B and D. SIMCA pattern recognition was, thereafter, used to find the "pattern" (data structure) in each of the two classes (stable and unstable). Finally, a test set of 109 complexes of unknown stability was classified by comparing the structural vector of each test-complex with the two "class patterns":

The rather low classification rate of 75% (estimated from the classification of the learning set) was probably due to several factors.

(i) A large structural diversity within both classes. Pattern recognition, like all methods based on empirical models, works best on data sets with a high degree of homogeneity in each class. A single class of stable hydrides ranging from $LiBH_4$ to $CsBH(NC_4H_4)_3$ and $KAl(iso-Bu)_3$ is likely to cause problems because so many structural properties vary widely in the class.

(ii) Difficulties with the definition of the properties "stable" and "unstable". In particular the unstable hydrides may be labeled unstable in the literature for several reasons: Difficulties in their synthesis, their instability in unfortunately selected solvents, and sheer bad luck in various stages of their preparation.

(iii) The choice of structural description was rather naive, in particular for the ligands D.

In the present paper we try to avoid these difficulties by (i) using separate models for different types of complexes, *i.e.* one model for ABH_4 , a second for ABH_3D *etc.* and (ii) allowing for the possibility of asymmetric data structures as recently discussed² and described below for the asymmetric analysis of "unsubstituted" complexes ABH_4 . In the following paper we shall report on complexes ABH_3D using also an extended set of descriptors for the structure of the ligands D. Studies on ABH_2D_2 and $ABHD_3$ are in progress.

THEORY

The philosophy of pattern recognition and the SIMCA methodology have been described in detail in part 1. In summary, pattern recognition is based on the description of objects (in the present case complexes) as M -dimensional data vectors. In the present case the elements in these vectors are derived from the chemical structures of the complexes.

The objects are divided into classes relating to the investigated problem (here stable or unstable). The pattern recognition method "learns" the "pattern"

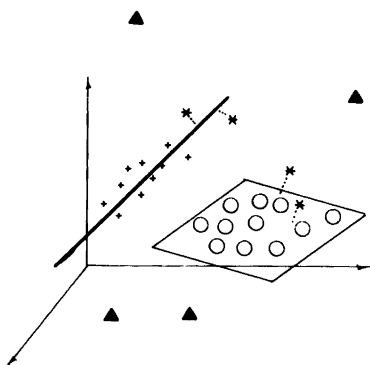


Fig. 1. A one-dimensional (line) and two-dimensional (plane) PC model in a three-dimensional M -space. Crosses and rings correspond to the training set of the two class models. Asterisks symbolize points sufficiently close to one of the models to be assigned to the corresponding class. Filled triangles symbolize outliers not similar to any of the classes.

of each class from objects with known class assignment (the training set) and then uses these patterns to classify new objects of initially unknown class assignment.

The SIMCA method is based on modelling each class training matrix by a separate principal components (PC) model (eqn. (1) in part 1). After the deletion of irrelevant variables, each data vector

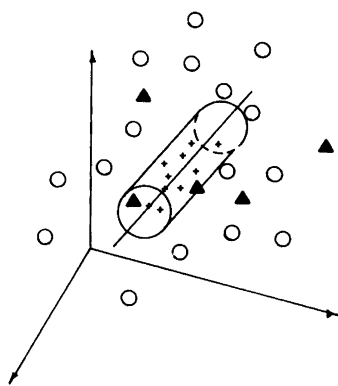


Fig. 2. A three dimensional M -space with two classes, one of which (crosses) is well described by a one term PC model (line), but the second (rings) having no structure. By means of the confidence interval around the class 1 model (cylinder), points in the test set (filled triangles) can still be classified according to their position inside or outside the confidence cylinder.

in the test set is fitted to each of the class PC models and assigned to the class it fits best. This can be illustrated in terms of points and linear models in an M -dimensional space defined by one coordinate axis for each of the M variables (Fig. 1). Recently we have found, however, that the classification problem often is asymmetric.² In such cases one of the classes still has structure and can be described by means of a PC model but another class has no systematic structure and cannot be modelled. This situation is shown in Fig. 2 for the simplified case of $M = 3$.

Table 1. Complexes ABH_4 analyzed. Complexes 1–15 belong to the training set of class 1. Column 4 shows the residual SD of the complexes when fitted to the resulting PC model of class 1. Column five shows the F value obtained by squaring the ratio between the SD in column 4 and the typical SD of class 1 (0.26). This F value should be compared with a critical F with 14 and 168 degrees of freedom ($F_{.05} = 1.94$).

	A	B	Initial assignm.	SD	F	
1	Li	B	1	0.14	0.31	
2	Na	B	1	0.20	0.62	
3	K	B	1	0.26	1.0	
4	Rb	B	1	0.10	0.15	
5	Cs	B	1	0.17	0.46	
6	Li	Al	1	0.27	1.1	
7	Na	Al	1	0.31	1.4	
8	K	Al	1	0.35	1.8	
9	Rb	Al	1	0.25	0.97	
10	Cs	Al	1	0.29	1.3	
11	Li	Ga	1	0.25	0.95	
12	Na	Ga	1	0.29	1.3	
13	K	Ga	1	0.33	1.7	
14	Rb	Ga	1	0.23	0.80	
15	Cs	Ga	1	0.27	1.1	
16	Li	In	0 (2)	0.52	4.1	$p < 0.01$
17	Li	Tl	0 (2)	0.98	15.0	
18	Na	In	0	0.53	4.3	
19	K	In	0	0.55	4.6	
20	Rb	In	0	0.51	3.9	
21	Cs	In	0	0.53	4.2	
22	Na	Tl	0	0.99	15	
23	K	Tl	0	1.00	15	
24	Rb	Tl	0	0.98	14	
25	Cs	Tl	0	0.99	15	

This mathematical inhomogeneity of the second class corresponds, chemically, to the members in this class not being chemically similar even if they have some property in common. This is the case when this property, say instability, can be caused by different independent "effects" – the class then, in fact, contains representatives of several inherently different groups.

In the present study of the unsubstituted complexes ABH_4 , we have only two complexes in the unstable class (Table 1). This is too few to allow the estimation of a class PC model and we, therefore, have to treat the data analysis as asymmetric – only class 1 (stable) can be modelled.

DATA

Of the 25 possible complexes with the formula ABH_4 , 15 have been reported as stable and 2 as unstable (Table 1).³ The 15 stable complexes are treated as a training set with one class (stable) and the unstable and unassigned complexes constitute the test set.

The same variables are used to describe the atoms A and B as in part 1, *i.e.* in all 16 variables for each complex. The eight variables for the atoms were: m.p., b.p., density, atomic, covalent and ionic radii, 1st ionization energy and electronegativity.

RESULTS

The data were first normalized (Table 2). Cross-validation⁴ then shows that two terms are significant in the PC model of class 1; *i.e.* the data points of complexes 1–15 lie close to a plane in the 16 dimensional M -space. The distances between all 25 complexes and this plane are shown in Table 1 (residual SD). The stable complexes have a typical SD of 0.26. The 2 unstable complexes lie significantly further away as do all the other unassigned complexes. Hence they are all predicted to be unstable. All 16 variables have a modelling power exceeding 0.45 and are, therefore, all considered to be relevant. The parameters α_i , β_{1i} and β_{2i} are shown in Table 2. Finally, Fig. 3 shows a three dimensional projection of the 16-dimensional M -space. The 25 possible complexes seem to lie on a parabolic-cylindrical surface and it is easily seen how a plane well fits the stable complexes 1–15 and at the same time passes far from the remaining complexes.

Validation of the results. Validation of the classification results was made as follows:

1. Class 1 (stable) was divided into 4 groups: The first containing compounds 1,5,9,13, the second containing 2,6,10,14, the third containing 3,7,11,15 and the fourth 4,8,12.

2. 4 separate PC analyses were made using as training set three quarters of the original training

Table 2. The data were initially normalized by the transformation $y_{ik}^* = (y_{ik} - y_i^-) / S_i$. The parameters of class 1 $\alpha_i^{(1)}$, $\beta_{1i}^{(1)}$ and $\beta_{2i}^{(1)}$ refer to the analysis of the normalized data y_{ik}^* . ψ_i is the corresponding modelling power⁵ of each variable based on class 1. ϕ_i is the discrimination power using the two component PC model for class 1 and the model $y_{ik}^* = \alpha_i^{(2)}$ for class 2.

	\bar{y}_i	S_i	$\alpha_i^{(1)}$	$\beta_{1i}^{(1)}$	$\beta_{2i}^{(1)}$	ψ_i	ϕ_i	$\alpha_i^{(2)}$
1	935	623	-.186	-.189	.289	.79	8.8	1.396
2	926	278	-.194	-.178	.273	.65	5.4	1.452
3	108	52.0	.142	.186	-.285	.50	3.1	-1.067
4	212	44.9	.168	.199	-.304	.91	20	-1.258
5	179	45.3	.164	.199	-.305	.88	14	-1.232
6	114	42.8	.168	.199	-.305	.95	36	-1.258
7	108	14.2	-.151	-.200	.306	.75	6.5	1.129
8	8.59	1.12	-.168	-.197	.300	.82	9.2	1.259
9	951	1030	.093	.338	.221	.82	6.8	-.701
10	2740	828	.163	.312	.204	.99	3.6	-1.219
11	435	263	-.265	-.138	-.090	.45	2.3	1.987
12	132	24.3	-.199	-.290	-.190	.85	9.7	1.493
13	113	22.3	-.197	-.294	-.192	.93	14	1.477
14	49.2	22.6	-.229	-.264	-.173	.84	5.0	1.716
15	154	25.0	.088	.341	.223	.88	5.5	-.659
16	17.1	2.11	-.028	.337	.221	.65	4.0	.209

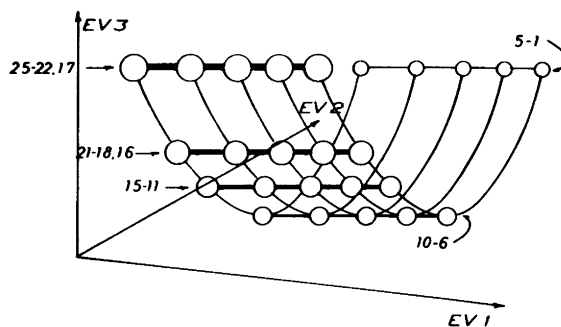


Fig. 3. Three-dimensional eigenvector-projection of the 16-dimensional M -space containing the 24 hydride data vectors.

set with on of the above groups made into an artificial test set with known expected classification. The data were scaled in the same way as in the full analysis (Table 2).

3. The deleted group was classified using the PC parameters from the analysis of the corresponding reduced training set. In addition, the two "unstable" complexes were fitted to the same model to verify the difference in fit between stable and unstable compounds. The result of the validation showed (as expected from Fig. 3) that each compound was classified correctly when being in one of the test sets. Stable compounds fitted the PC models with a typical RSD of 0.26 with a range from 0.11 to 0.33. The two unstable compounds 16 and 17 were never closer to class 1 than 0.49 and 0.93 RSD, respectively. These results are in perfect agreement with those shown in Table 1 indicating a very stable classification structure.

DISCUSSION

The complexity of the data analysis is substantially reduced by analyzing the unsubstituted complexes ABH_4 separately. This separate analysis does not decrease the information content of the results, however, since the knowledge about a complex being ABH_4 , ABH_3D , ABH_2D_2 or $ABHD_3$ exists *a priori* also for "non-made" complexes.

The separate analysis also simplifies the interpretation of the results. The positions of the complexes on the three dimensional surface depicted in Fig. 3 indicate that the direction parallel to the second eigenvector (EV 2) is the main factor related to the stability of the complexes ABH_4 . Little information presently exists on the variation of

stability along the first (EV1) or third (EV3) eigenvectors. We presently can only conclude that as long as no major change in stability is related to the latter eigenvectors, the complexes 16–25 are all classified as unstable. This further underlines the dramatic increase in information that is obtained when a quantitative variable has been measured which is related to the problem of classification.² In the present case, this quantitative variable would involve the *degree* of stability of the complexes. A quantitative knowledge about such a variable would allow a level three PARC analysis² which would relate the degree of stability to the position in the "stable" class. Such analysis would give information about the importance of the direction EV1 and EV3 and thus considerably sharpen the predictions for the "test set" complexes 18–25.

The three eigenvectors in Fig. 3 describe 95 % of the variance in the 16 variables of atoms A and B. Hence the three eigenvectors can be taken as new coordinates containing almost the same information as the original 16 variables. One of the eigenvectors relates only to atom A and the other two to atom B. Table 3 shows the new coordinates on an arbitrarily

Table 3. Variables summarizing the original 16 variables of atoms A and B. The variables Z are scaled to be 0.0 for Li and Al.

Atom A	Z_1	Atom B	Z_2	Z_3
Li	0.0	B	8.8	9.0
Na	6.0	Al	0.0	0.0
K	8.5	Ga	-3.0	1.8
Rb	10.5	In	-3.8	6.0
Cs	11.2	Tl	-5.5	16.0

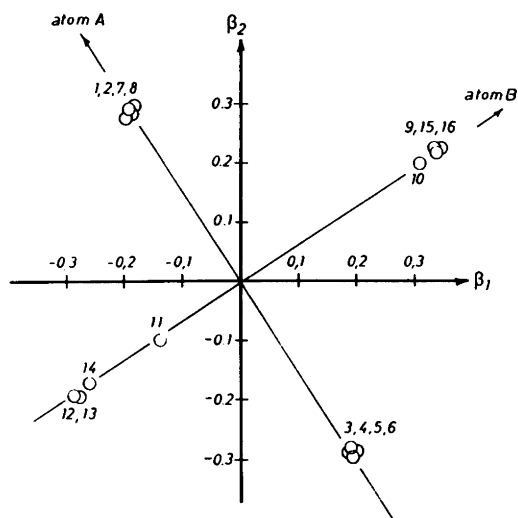


Fig. 4. Plot of the two parameter vectors β_1 against β_2 (class 1) showing the grouping of the variables 1 through 16.

chosen scale. This finding is in accordance with recent experimental results.⁶

The interpretation of the class parameters β for the "stable" class (Table 2) is fairly straightforward. Fig. 4 shows a plot of β_2 against β_1 which reveals two groups of variables each for atoms A and B; one containing variables increasing with increase in atom number and the other containing variables with the opposite behaviour. The figure also shows that a rotation to two orthogonal factors, one for atom A and the other for atom B retains all information.

Acknowledgement. The project was supported by a grant from the Swedish Natural Science Research Council.

REFERENCES

1. Strouf, O. and Wold, S. *Acta Chem. Scand. A* 31 (1977) 391.
2. Albano, C., Dunn, W., III, Edlund, U., Johansson, E., Nordén, B., Sjöström, M. and Wold, S. *Anal. Chim. Acta. Comput. Optim.* 103 (1978) 429.
3. Wold, S. and Strouf, O. *Data Base II, Stability of Complex Hydrides*; Tech. report, December 1978, Research Group for Chemometrics, Institute of Chemistry, Umeå University, Umeå 1978.

4. Wold, S. *Technometrics* 20 (1978) 397.
5. Wold, S. and Sjöström, M. In Kowalski, B. R., Ed., *Chemometrics, Theory and Application, ACS Symposium Ser. 52*, Am. Chem. Soc., 1977.
6. Strouf, O. and Fusek, J. *Collect. Czech. Chem. Commun.* (1979). In press.

Received February 19, 1979.