A Nonlinear Map of Substituent Constants for Selecting Test Series and Deriving Structure-Activity Relationships. 2. Aliphatic Series

Daniel Domine,[†,‡] James Devillers,[*,†] and Maurice Chastrette[‡]

CTIS, 21 rue de la Bannière, 69003 Lyon, and Laboratoire de Chimie Organique Physique, U.R.A. CNRS 463, Université Lyon-I, 43 Bd du 11 Novembre 1918, 69622 Villeurbanne CEDEX, France

Received August 23, 1993[•]

A nonlinear mapping (NLM) analysis was performed on a set of 103 aliphatic substituents described by five variables encoding hydrophobic (Fr), steric (MR), and electronic effects (HBA, HBD, and *F*). NLM allowed to easily summarize the main information contained in the original data table. By means of collections of graphs, it was possible to relate the structure of the aliphatic substituents to their Fr, MR, HBA, HBD, and *F* values. The proposed approach provides a useful and easy tool for the selection of test series and for deriving structure-activity relationships.

Introduction

Quantitative structure-activity relationship (QSAR) studies are based on the modification of the structure of known series of active chemicals and the modeling of the effects of these changes on their biological activities.[1,2] It is well known that the physicochemical properties of molecules are the basic parameters to be taken into account in these kinds of studies, and as a result, numerous data compilations of substituent constants have been elaborated for use in QSAR studies.[2-5] Among them, the most widely used are the π contribution of Hansch which depicts the lipophilic character of the substituents, the Hammett σ constants which are used to account for electronic processes, the Swain and Lupton *F* and *R* parameters derived from the σ constants separating the inductive and resonance effects of the substituents, and the molar refractivity (MR) used to describe the steric bulk of substituents.[6-11]

In drug design, it is essential to select test series with high information content in order to reduce the costs in research by maximizing the information content obtained from each molecular probe in a set of congeners. A lot of works have been directed toward this aim, and numerous methods have been proposed.[12-28] Significant advances dealt with the use of a multivariate method[14] and later of a graphical representation of the data allowing selection of test series by simple visual inspection of a 2-D map summarizing the information content of a matrix of physicochemical properties.[20-22,28] For a comprehensive review, one should refer to the paper of Pleiss and Unger.[29] Even if these methods have been successfully used, it must be pointed out that, from a practical point of view, none of these approaches are completely satisfactory.[30] In this context, we recently proposed the use of an original graphical approach based on the nonlinear mapping (NLM) method.[30] We showed that it was possible to obtain an easily interpretable nonlinear map of aromatic substituent constants for the selection of test series and the derivation of structure-activity relationships (SAR). The good results obtained prompted us to perform the same kind of analysis on a set of aliphatic substituents.

Nonlinear Mapping

The nonlinear mapping (NLM) method was designed by Sammon[31] and introduced in chemistry by Kowalski and Bender.[32,33] It is based on a concept similar to multidimensional scaling (MDS)[34-36] and is aimed at representing a set of points defined in an *n*-dimensional space by a human-perceivable configuration of the data in a lower *d*-dimensional space ($d = 2$ or 3). NLM tries to preserve distances between points in the display space as similar as possible to the actual distances in the original space. The procedure for performing this transformation is summarized in Figure 1. Briefly, it consists in calculating a mapping error (*E*) between the distances in the original space and the distances in the display space (i.e., nonlinear map). This error is used to modify the coordinates of points in the display space. This process is carried out iteratively by means of a minimization algorithm called "steepest descent procedure" until termination conditions are satisfied. The most widely used termination condition is a sufficiently low difference between the error calculated at step *n* and step *n*-1 in the iteration process.

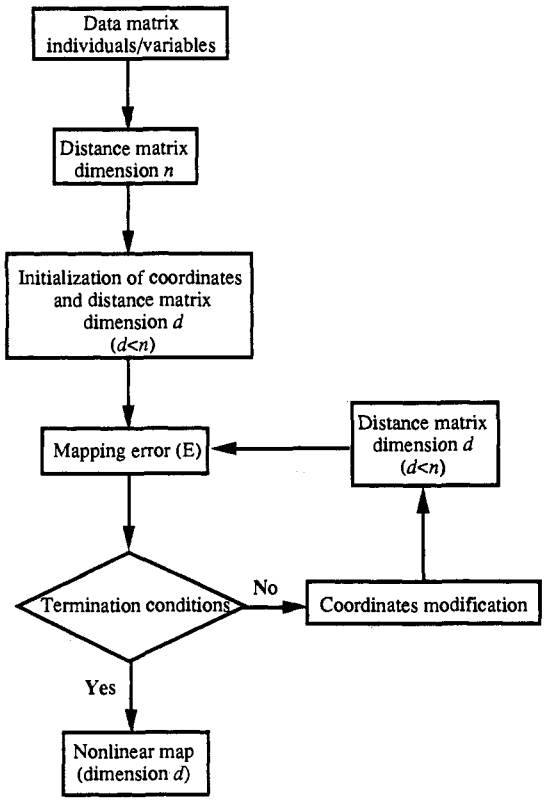

Figure 1. NLM algorithm flow diagram.

* Author to whom all correspondence should be addressed.
† CTIS.
‡ Université Lyon-I.
• Abstract published in *Advance ACS Abstracts*, March 1, 1994.

0022-2623/94/1837-0981$04.50/0

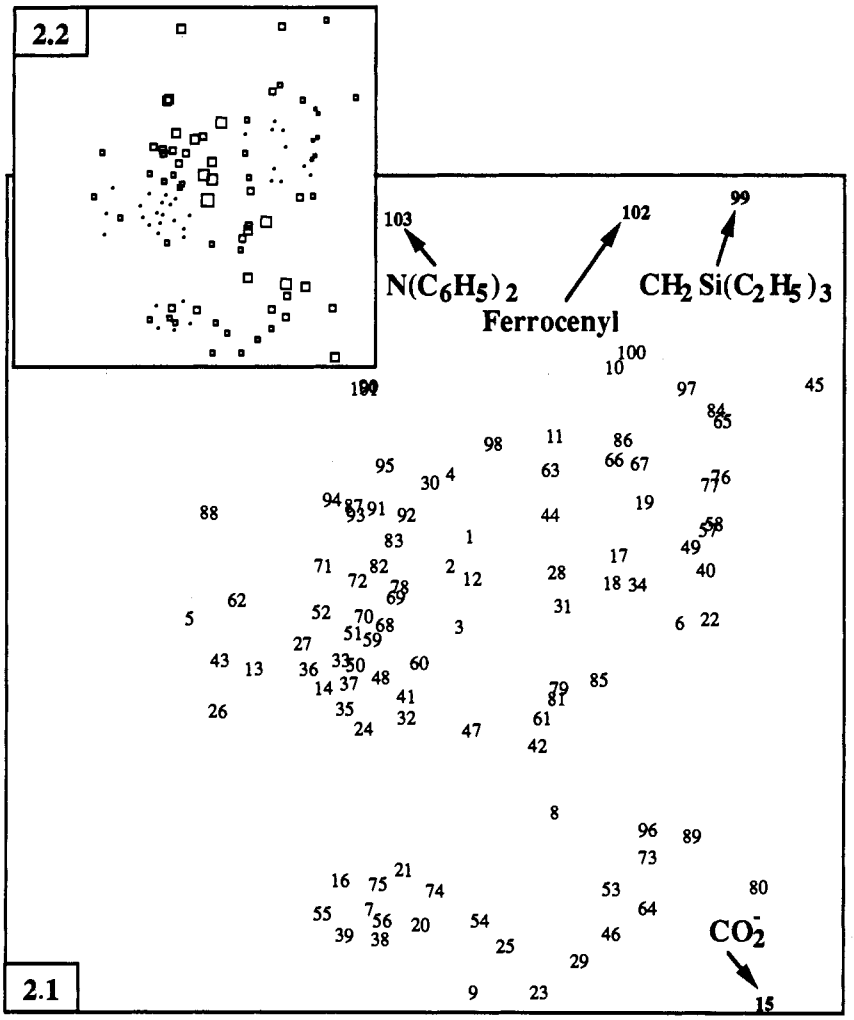

Figure 2. (2.1) Nonlinear map of the 103 aliphatic substituents described by five substituent constants (Fr, HBA, HBD, MR, and *F*). (2.2) Plot of the individual mapping errors on each substituent of the nonlinear map. Squares are proportional in size to the magnitude of the errors. 1, Br; 2, Cl; 3, F; 4, I; 5, NO_2; 6, H; 7, OH; 8, SH; 9, NH_2; 10, CBr_3; 11, CCl_3; 12, CF_3; 13, CN; 14, SCN; 15, CO_2^-; 16, CO_2H; 17, CH_2Br; 18, CH_2Cl; 19, CH_2I; 20, $CONH_2$; 21, CH=NOH; 22, CH_3; 23, $NHCONH_2$; 24, OCH_3; 25, CH_2OH; 26, $SOCH_3$; 27, OSO_2CH_3; 28, SCH_3; 29, $NHCH_3$; 30, CF_2CF_3; 31: C≡CH; 32, CH_2CN; 33, $CH=CHNO_2$ (trans); 34, $CH=CH_2$; 35, $COCH_3$; 36, $OCOCH_3$; 37, CO_2CH_3; 38, $NHCOCH_3$; 39, C=O(NHCH$_3$); 40, CH_2CH_3; 41, OCH_2CH_3; 42, CH_2OCH_3; 43, SOC_2H_5; 44, SC_2H_5; 45, $CH_2Si(CH_3)_3$; 46, NHC_2H_5; 47, $N(CH_3)_2$; 48, CH=CHCN; 49, cyclopropyl; 50, COC_2H_5; 51, $CO_2C_2H_5$; 52, $OCOC_2H_5$; 53, $EtCO_2H$; 54, $NHCO_2C_2H_5$; 55, $CONHC_2H_5$; 56, $NHCOC_2H_5$; 57, $CH(CH_3)_2$; 58, C_3H_7; 59, $OCH(CH_3)_2$; 60, OC_3H_7; 61, $CH_2OC_2H_5$; 62, SOC_3H_7; 63, SC_3H_7; 64, NHC_3H_7; 65, $Si(CH_3)_3$; 66, 2-thienyl; 67, 3-thienyl; 68, $CH=CHCOCH_3$; 69, $CH=CHCO_2CH_3$; 70, COC_3H_7; 71, $OCOC_3H_7$; 72, $CO_2C_3H_7$; 73, $(CH_2)_3CO_2H$; 74, $NHCOC_3H_7$; 75, $CONHC_3H_7$; 76, C_4H_9; 77, $C(CH_3)_3$; 78, OC_4H_9; 79, $CH_2OC_3H_7$; 80, NHC_4H_9; 81, $N(C_2H_5)_2$; 82, $CH=CHCOC_2H_5$; 83, $CH=CHCO_2C_2H_5$; 84, C_5H_{11}; 85, $CH_2OC_4H_9$; 86, C_6H_5; 87, OC_6H_5; 88, $SO_2C_6H_5$; 89, NHC_6H_5; 90, 2-benzthiazolyl; 91, $CH=CHCOC_3H_7$; 92, $CH=CHCO_2C_3H_7$; 93, COC_6H_5; 94, $CO_2C_6H_5$; 95, $OCOC_6H_5$; 96, $NHCOC_6H_5$; 97, $CH_2C_6H_5$; 98, $CH_2OC_6H_5$; 99, $CH_2Si(C_2H_5)_3$; 100, $CH=CHC_6H_5$ (trans); 101, $CH=CHCOC_6H_5$; 102, ferrocenyl; 103, $N(C_6H_5)_2$.

Experimental Section

A nonlinear mapping analysis[31,37] was performed on a set of 103 aliphatic substituents[4] described by means of five substituent constants encoding their hydrophobic, steric, and electronic effects. These parameters were respectively the hydrophobic constant for aliphatic substituents Fr,[4] the molar refractivity (MR), the H-bonding acceptor (HBA) and donor (HBD) abilities, and the inductive parameter *F* of Swain and Lupton.[9] Data were obtained from a compilation of Hansch and Leo.[4] It is obvious that quantitative values for the H-bonding abilities[38,39] would have been better, but available data were still too scarce to handle all the substituents of our data set. For the NLM analysis, Fr, MR, and *F* were centered (i.e., zero mean) and reduced (i.e., unit variance). HBA and HBD were not centered and the 1's were replaced by a value yielding a unit variance. Note that, due to the fact that calculations were performed on the distance matrix, the results were not affected by the centering. It was only performed to improve the visualization of the data when they were reported on the map. The results obtained were interpreted in terms of structure–property and structure–activity relationships (SPR and SAR) by plotting various qualitative and quantitative information on the nonlinear map derived as recently described.[37,40-43] Before interpreting the map, its statistical significance was verified by inspecting the total mapping error and the goodness of fit of each point.[30,37] The NLM analysis was performed with the STATQSAR package[44] and the graphical analysis with GraphMu.[45]

Results and Discussion

Figure 2.1 shows the nonlinear map of the 103 aliphatic substituents described by five substituent constants (i.e., Fr, HBA, HBD, MR, and *F*). It is worth noting that principal components analysis on standardized data (i.e., unit variance and zero mean) did not allow to reduce the dimensionality of the data matrix in two dimensions since the first two factors only explained 68% of the total variance. At the opposite, with a low mapping error of $5.1e-2$ obtained with the NLM analysis, we can advance that the main information included in the matrix is

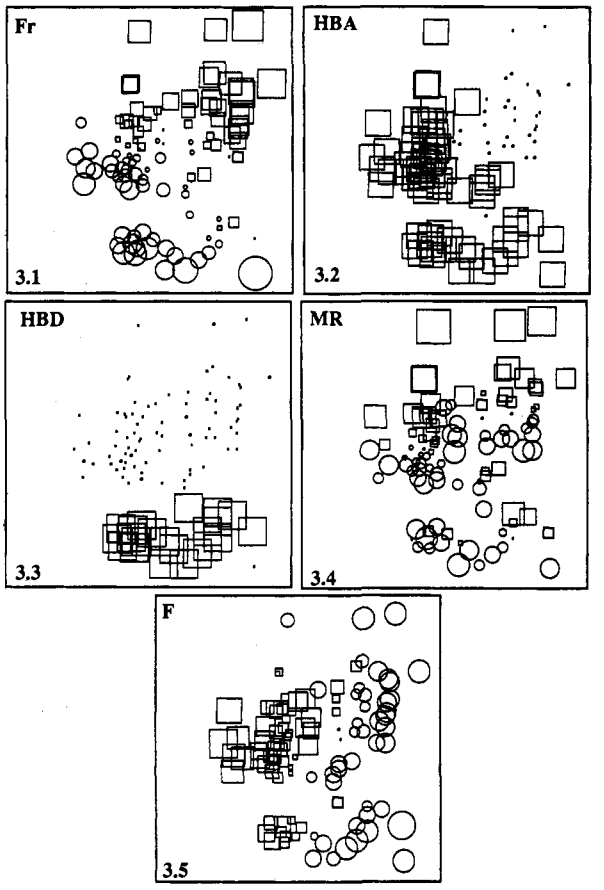

Figure 3. Plot of the scaled values of the five parameters on each substituent of the nonlinear map. Squares (positive values) and circles (negative values) are proportional in size to the magnitude of the parameters. In Figure 3.2,.3, the dots indicate the substituents which do not have the ability to accept and donate H-bonds, respectively.

summarized on the nonlinear map (Figure 2.1).[37] When the individual mapping error[37] of each point is plotted on the general map (Figure 2.1) by means of squares proportional to the magnitude of the individual errors, it appears (Figure 2.2) that the points located in the middle of the cloud are not as well represented as the other substituents. However, due to the low magnitude of the total mapping error, Figure 2.1 can be interpreted without paying special attention to any particular point. A rapid inspection of this figure underlines the particular position of substituents $n°$ 15 (CO_2^-), 99 ($CH_2Si(C_2H_5)_3$), 102 (ferrocenyl), and 103 ($N(C_6H_5)_2$). Substituent $n°$ 15 bears a charge and therefore possesses particular constants. It has the lowest Fr value of the set. The remaining substituents are characterized by their high MR values.

I. Interpretation of the Nonlinear Map in Terms of Structure–Property Relationships. A. Representation of the Data on the Nonlinear Map. Figure 2.1 could be directly interpreted in terms of SPR, but this would require constant reference to the original data and to the structural features of the substituents. To facilitate this work, the original values for each parameter have been plotted on the nonlinear map by means of squares (positive values) and circles (negative values) whose sizes are proportional to the magnitude of the studied parameters. Values represented are those used for the NLM analysis. Briefly, the larger the square, the larger the value, and the larger the circle, the smaller the value. Figure 3 shows that the substituents are distributed and clustered on the nonlinear map according to the values of Fr, HBA, HBD, MR, and *F*. Indeed, for all parameters, gradients or clusters can be observed. Thus, in Figure 3.1 Fr values of the substituents increase from the bottom to the top of the map. In the same way, there are obvious clusterings of HBA and HBD substituents in Figure 3.2,.3. Figure 3.4 shows that MR values of substituents increase from the bottom to the top of the map with some exceptions. Last, Figure 3.5 reveals an increasing gradient of the *F* values running from the right-hand side to the left-hand side of the map.

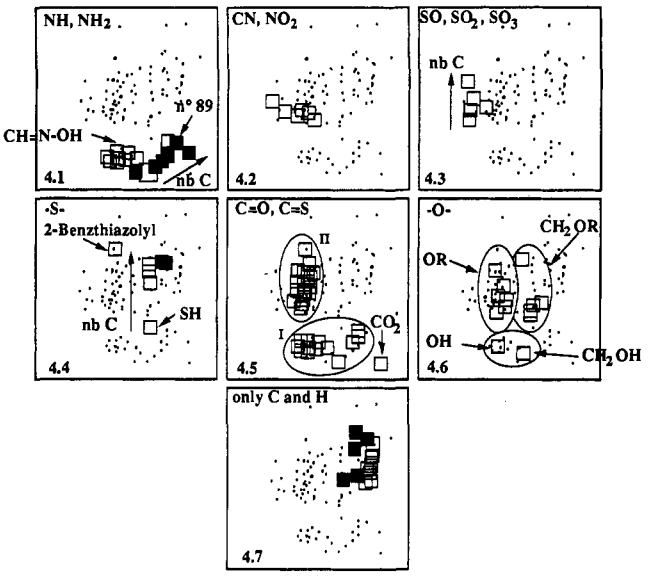

Figure 4. Plot of the presence or frequency of some functional groups or skeleton similarities in the aliphatic substituents. Squares are proportional in size to the number of groups. The absence of a functional group is represented by a dot.

B. Projection on the Nonlinear Map of the Presence or Frequency of Functional Groups. In order to give a full description of the nonlinear map in terms of structural information and underline SPR for the 103 aliphatic substituents under study, various structural elements have been reported on the nonlinear map (Figure 2.1). Thus, a series of figures was drawn on which was represented the presence or frequency of some functional groups and/or skeleton similarities. Figure 4 shows the relative positions of the clusters as well as the coherence of the map from a chemical point of view. Figure 4.1 shows that substituents containing primary or secondary amine groups are all located at the bottom of the map. Comparison with Figure 3 reveals that these chemicals are characterized by low Fr values (Figure 3.1) and HBA and HBD abilities (Figure 3.2,.3). The only exception is substituent $n°$ 89 (NHC_6H_5) which has a higher value of Fr. A closer inspection of Figure 4.1 stresses that substituents with the general formula NHR with R = H, alkyl, or phenyl ($n°$ 9, 29, 46, 64, 80, and 89, filled in black in Figure 4.1) are located on the right-hand side of the cloud of squares while the other substituents containing a carbonyl group are preferentially located on the left side ($n°$ 20, 38, 39, 54, 55, 56, 74, and 75) except substituents $n°$ 23 and 96 (i.e., $NHCONH_2$ and $NHCOC_6H_5$). Figure 3.5 shows that this separation principally results from differences in their *F* values. Last, among the former cluster (Figure 4.1), it is noteworthy that a gradient depending on the number of carbon atoms can be observed. On this figure, substituent $n°$ 21 (CH=NOH), which presents the same properties as the substituents of the second cluster, has also been represented.

In Figure 4.2, we have represented the number of CN and NO_2. All the substituents containing these groups ($n°$ 5, 13, 14, 32, 33, and 48) are located in the same region of the nonlinear map. Comparison of Figure 4.2 with Figure 3.1 shows that this location can be explained by the low Fr values. In addition, these substituents have the ability to accept but not to donate H-bonds (Figure 3.2,.3).

Figure 4.3 shows that substituents containing SO, SO_2, or SO_3 groups ($n°$ 26, 27, 43, 62, and 88) are also preferentially located in the left-hand side of the cloud of points on the map, indicating that their Fr values are generally low (Figure 3.1) and that they have the ability to accept H-bonds (Figure 3.2). It is noteworthy that they have rather similar properties to cyano and nitro groups, since they are located in the same region of the map. A closer inspection of this figure reveals a gradient linked to the number of carbon atoms of the alkyl group (nb C in Figure 4.3).

Figure 4.4 shows that the substituents containing an -S- group form a cluster on the map. The two outliers observed are substituents $n°$ 8 and 90. The former is the SH group which is characterized by its ability to donate H-bonds (Figure 3.3). The latter is 2-benzthiazolyl which is very particular due to its large MR value (Figure 3.4) and its HBA ability (Figure 3.2). Among the cluster stressed above, one can separate the cyclic substituents $n°$ 66 and 67 (i.e., 2- and 3-thienyl) on the right-hand side (indicated in black on Figure 4.4) from those having the general formula -SR with R = alkyl ($n°$ 28, 44, and 63) on the left-hand side for which a gradient depending on the number of carbon atoms can be drawn.

On Figure 4.5 has been plotted the presence of C=O or C=S groups. These substituents form two clusters. The substituents of the bottom cluster (cluster I) are characterized by HBA and HBD abilities (Figure 3.2,.3), while those of the top cluster (cluster II) can only accept H-bonds. Only substituent $n°$ 15 (CO_2^-), whose atypical location was stressed earlier in this paper, transgresses this rule. A more detailed inspection of this plot could reveal gradients depending on the number of carbon atoms for each chemical family running in the vertical direction to the top of the map. Thus, for example, if we consider the esters with the general formula CO_2R with R = CH_3 ($n°$ 37), C_2H_5 ($n°$ 51), C_3H_7 ($n°$ 72), and C_6H_5 ($n°$ 94), it can be seen on Figure 2.1 that they are distributed along an axis linked to the number of carbon atoms they contain.

Inspection of Figure 4.6 clearly shows that the substituents containing an -O- group can have very different substituent constants according to the functional group in which they are involved but, also, their position (i.e., bound to the substitution site or not). Note that acids and esters have not been represented on this figure since they are included in Figure 4.5. A close inspection of Figure 4.6 reveals that three clusters can be isolated. First, at the bottom of the figure are found the substituents containing an alcohol group ($n°$ 7 and 25). They can donate H-bonds (Figure 3.3) and have lower Fr values (Figure 3.1). The group OH in which the oxygen atom is bound to the substitution site has a higher F value than the CH_2-OH group (Figure 3.5). The two remaining clusters concern ether groups. The first, located on the left-hand side, contains substituents with the oxygen atom bound to the substitution site. The second cluster consists of the substituents having the general formula CH_2OR. The substituents of the former cluster differ from those of the second by higher F values (Figure 3.5). For both clusters, gradients depending on the number of carbon atoms can be observed.

Figure 4.7 shows that substituents containing only C and H atoms cluster in the right-hand side of the map. These substituents are characterized by low F values (Figure 3.5). Furthermore, a closer inspection of Figure 4.7 reveals that a gradient depending on the number of carbon atoms can be drawn up. This gradient results in increased Fr and MR values (Figure 3.1,.4). It can also be noticed that the unsaturated substituents ($n°$ 31, 34, 86, 97, and 100, indicated in black) are located on the left side of the cloud of squares while H ($n°$ 6), cyclopropyl ($n°$ 49), and the alkanes ($n°$ 22, 40, 57, 58, 76, 77, and 84) are located on the right-hand side. Inspection of Figure 3.5 reveals that this results from slightly higher F values for the unsaturated substituents except for substituent $n°$ 97.

II. Selection of Test Series. Because the information contained in the large data table of 103 aliphatic substituents is broadly summarized on Figure 2.1 and also because Figures 3 and 4 give a full description of the nonlinear map in terms of structural information, it becomes very easy to inspect and compare the various possible substituents for SAR purposes. As stressed for the aromatic substituents in a previous paper,[30] the problem of the selection of test series is of the utmost importance. NLM can facilitate the selection of substituents for the constitution of nonredundant test series. This can simply be done by inspection of the map (Figure 2.1) by eye. Since it is possible to obtain different test series on the same map, synthetic feasibility or previous knowledge on the considered activity can always adequately be taken into account. It provides a very simple and straightforward representation of the results in line with classical chemical thinking, which should be attractive to synthetic chemists. Furthermore, the 2-D map gives a full picture of the data structure in the starting population which cannot be obtained with other classical multivariate methods.[4,14] The within-cluster position of substituents is now known, and no *a priori* decision as in other methods (except for the parameter space to be considered) is necessary to establish that map.

For comparison purposes, the results of a hierarchical cluster analysis (HCA) on the same data set[4] (Table 1) and those obtained by the same authors for the selection of a possible "ideal" test series of aromatic substituents[4] have been used. The substituents selected for the aromatic set which were present in the set of aliphatic substituents were kept and are represented on the nonlinear map (Figure 5) by means of squares. The black circle for NO_2 ($n°$ 5) indicates that this substituent has been replaced by CN ($n°$ 13, white square) for the aromatic substituents since NO_2 is readily reduced in biological systems.[4] The eight substituents indicated by white squares (Figure 5) are included in 6 of the 10 clusters derived by HCA[4] (Table 1). Indeed, OCH_3 ($n°$ 24) is located in cluster $n°$ 2 like CN ($n°$ 13). C_4H_9 ($n°$ 76) is located in cluster $n°$ 3 like H ($n°$ 6). From the HCA results, without the judgement of a chemist, it seems surprising to select two substituents in clusters $n°$ 2 and 3, while from this kind of approach, it would be more logical to only select one substituent per cluster. Inspection of our nonlinear map (Figure 5) shows that this choice is reasonable due to their differences and

Table 1. Aliphatic Substituent Constants-10 Cluster Sets[4,a]

cluster	n	members
1	6	**Br**, Cl, CF_3, F, I, CF_2CF_3
2	27	**NO_2**, CN, $SOCH_3$, SOC_2H_5, SOC_6H_7, SCN, CO_2CH_3, $OCOCH_3$, CH=$CHNO_2$ (trans), COC_2H_5, CH=CHCN, OSO_2CH_3, **OCH_3**, $COCH_3$, CH_2CN, OCH_2CH_3, $CO_2C_2H_5$, $OCH(CH_3)_2$, CH=$CHCOCH_3$, COC_3H_7, OC_3H_7, $OCOC_2H_5$, $OCOC_3H_7$, CH=$CHCO_2CH_3$, OC_4H_9, $CO_2C_3H_7$, CH=$CHCOC_2H_5$
3	20	**H**, CH_3, CH_2Br, CH_2Cl, CH=CH_2, SCH_3, CH_2CH_3, cyclopropyl, $CH(CH_3)_2$, C_3H_7, **C_4H_9**, $C(CH_3)_3$, CBr_3, CCl_3, SC_2H_5, SC_3H_7, CH_2I, 2-thienyl, C_6H_5, 3-thienyl
4	11	**OH**, CO_2H, CH=NOH, $CONH_2$, $NHCOCH_3$, C=O($NHCH_3$), $NHCO_2C_2H_5$, $CONHC_2H_5$, $NHCOC_2H_5$, $NHCOC_3H_7$, $CONHC_3H_7$
5	2	**SH**, C≡CH
6	11	NH_2, CH_2OH, $NHCH_3$, NHC_2H_5, $NHCONH_2$, $EtCO_2H$, NHC_3H_7, $(CH_2)_3CO_2H$, NHC_4H_9, NHC_6H_5, **$NHCOC_6H_5$**
7	7	CH_2OCH_3, $CH_2OC_2H_5$, $N(CH_3)_2$, $CH_2OC_6H_7$, **$N(C_2H_5)_2$**, $CH_2OC_4H_9$, $CH_2OC_6H_5$
8	7	$CH_2Si(CH_3)_3$, $Si(CH_3)_3$, C_8H_{11}, $CH_2C_6H_5$, **CH=CHC_6H_5 (trans)**, $CH_2Si(C_2H_5)_3$, ferrocenyl
9	11	CH=$CHCO_2C_2H_5$, CH=$CHCO_2C_3H_7$, $OCOC_6H_5$, **OC_6H_5**, CH=$CHCOC_3H_7$, COC_6H_5, $CO_2C_6H_5$, $SO_2C_6H_5$, 2-benzthiazolyl, CH=$CHCOC_6H_5$, $N(C_6H_5)_2$
10	1	CO_2^-

[a] The selected substituents (Figure 5) are indicated in bold. Note that the other cluster sets (Table VI-6a, p 60, and Table VI-6c, p 62)[4] could also have been used.

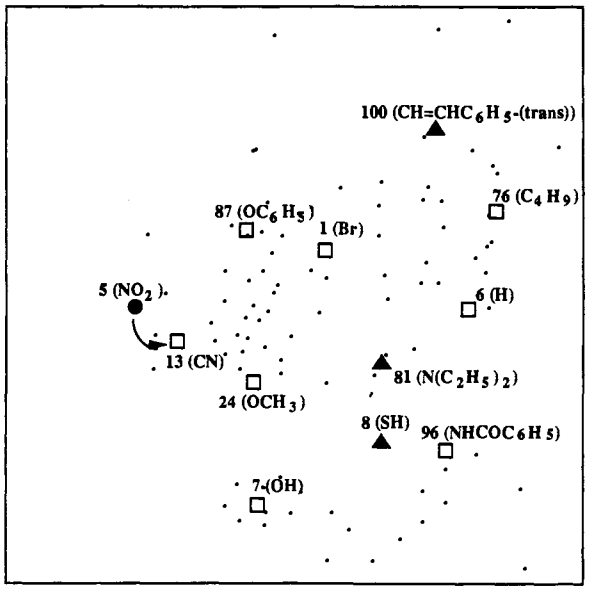

Figure 5. Test series selection.[4] For captions, see text.

their locations on the map. Substituents $n°$ 1 (Br), 7 (OH), 87 (OC_6H_5), and 96 ($NHCOC_6H_5$) are located in clusters $n°$ 1, 4, 9, and 6, respectively (Table 1). To complete the representation of the HCA results, it is necessary to select one substituent in clusters $n°$ 5, 7, 8, and 10 (Table 1). They are represented by black triangles in Figure 5.

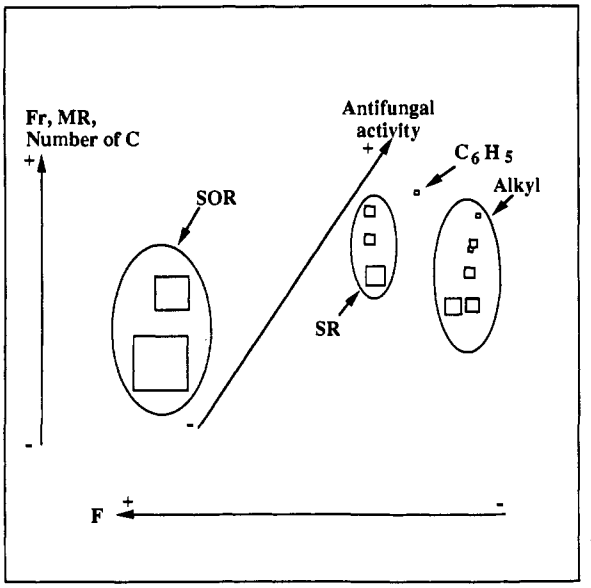

Figure 6. Plot of the antifungal activity of 12 1-(3,5-dichlorophenyl)-3-substituted-2,5-pyrrolidinediones against *B. cinerea*. Squares are proportional to $10^{-pI_{50}}$ values (mol L^{-1}) inducing 50% inhibition of mycelial growth.[46]

Cluster $n°$ 10 contains only CO_2^- which bears a charge. This substituent will be hardly comparable to the other substituents which are all uncharged, and we therefore prefer not to select it. To represent the three remaining clusters, we can propose, for example, SH ($n°$ 8) for cluster $n°$ 5, $N(C_2H_5)_2$ ($n°$ 81) for cluster $n°$ 7, and CH=CHC_6H_5 (trans) ($n°$ 100) for cluster $n°$ 8. Note that, from Table 1, other combinations are possible depending on the goal of the study.

Representative substituents of each HCA cluster (Table 1) cover a broad range of substituent constants as they appear widely spread on the nonlinear map (Figure 5). In addition, the nonlinear map allows visualization of the selection and the relative locations of the substituents within and between clusters. However, it must be noted that the selection of a test series cannot be universal and always depends on the studied activity and the knowledge we have on it.

III. Deriving Structure–Activity Relationships from the Nonlinear Map. As shown in a previous study on aromatic substituents,[30] our graphical approach allows one to represent on the map the biological activities and to underline relationships between structures and biological responses. For illustrative purposes, an example dealing with the antifungal activity (i.e., $pI_{50} = \log(1/I_{50})$ where I_{50} values express the molar concentration for 50% inhibition of mycelial growth) of 1-(3,5-dichlorophenyl)-3-substituted-2,5-pyrrolidinediones against *Botrytis cinerea* is presented.[46] From a set of 22 pI_{50} values,[46] we only kept the 12 values corresponding to substituents contained in our set. It must be noted that the remaining 10 substituents did not carry much more information and that the 12 selected substituents are representative of the set. These values were transformed as $10^{-pI_{50}}$ for a better visualization of the relationships between the substituents and the biological activity. Therefore, the smaller the square, the larger the antifungal activity. Figure 6 stresses two gradients. The former runs from the right-hand side to the left-hand side of the map. It is associated with variation in F values (Figure 3.5). The latter runs from the bottom to the top of the map and is associated with

variations in Fr (Figure 3.1) and MR (Figure 3.4) data. A closer inspection of this map reveals that the gradient of F values corresponds to three different functional groups. From the right-hand side to the left-hand side are found the alkyl groups, the thioethers, and, last, the S=O groups. Then, within these clusters, the variations of Fr and MR are associated with an increase in the number of carbon atoms. These results confirm those published in the literature[46] which showed, from classical regression equations, a relationship between the antifungal activity of 1-(3,5-dichlorophenyl)-3-sub-stituted-2,5-pyrrolidinediones and the log P. A new examination of these data revealed that both the size of the substituents and their dipole moment were also important.[47]

Our results confirm those obtained with the aromatic substituent constants.[30] They show that NLM coupled to graphical tools summarizes the main part of the information contained in the large aliphatic substituent constants data table of Hansch and Leo.[4] They also provide a useful and easy tool for the selection of test series. Indeed, to ensure a broad spectrum of substituent constants, it suffices to select substituents in the different regions of the nonlinear map. This method is open, and any information susceptible to help in the selection of the data set (e.g., synthetic feasibility, previous knowledge on the biological activity) can be plotted on the nonlinear map. For particular cases such as multiple substitutions or larger differences in structures of the chemicals to be studied, it may be useful to recalculate a nonlinear map with the variables suspected to influence the activity under study and to perform a new selection as described in this paper. Last, our results clearly demonstrate that from the nonlinear map (Figure 2.1) it is very easy to derive SAR.

References

(1) Devillers, J. Statistical Analyses in Drug Design and Environmental Chemistry: Basic Concepts. In *Quantitative Structure/Activity Relationships (QSAR) in Toxicology*, Coccini, T., Giannoni, L., Karcher, W., Manzo, L., Roi, R., Eds.; Commission of the European Communities: Luxembourg, 1992; pp 27–41.

(2) Norrington, F. E.; Hyde, R. M.; Williams, S. G.; Wootton, R. Physicochemical-Activity Relations in Practice. 1. A Rational and Self-Consistent Data Bank. *J. Med. Chem.* **1975**, *18*, 604–607.

(3) Hansch, C.; Leo, A.; Unger, S. H.; Kim, K. H.; Nikaitani, D.; Lien, E. J. "Aromatic" Substituent Constants for Structure-Activity Correlations. *J. Med. Chem.* **1973**, *16*, 1207–1216.

(4) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; John Wiley & Sons: New York, 1979.

(5) Hansch, C.; Leo, A.; Taft, R. W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem. Rev.* **1991**, *91*, 165–195.

(6) Hansch, C., Fujita, T. ρ-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616–1626.

(7) Bowden, K. Electronic Effects in Drugs. In *Comprehensive Medicinal Chemistry*; Ramsden, C. A., Ed.; Pergamon Press: Oxford, 1990; Vol. 4, pp 205–239.

(8) Dearden, J. C. Physico-Chemical Descriptors. In *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W., Devillers, J., Eds.; Kluwer Academic Publishers: Dordrecht, 1990; pp 25–59.

(9) Swain, C. G.; Lupton, E. C. Field and Resonance Components of Substituent Effects. *J. Am. Chem. Soc.* **1968**, *90*, 4328–4337.

(10) Dunn, W. J. Molar Refractivity as an Independent Variable in Quantitative Structure-Activity Studies. *Eur. J. Med. Chem.* **1977**, *12*, 109–112.

(11) Livingstone, D. J. Quantitative Structure-Activity Relationships. In *Similarity Models in Organic Chemistry, Biochemistry and Related Fields*; Zalewski, R. I., Krygowski, T. M., Shorter, J., Eds.; Elsevier: Amsterdam, 1991; pp 557–627.

(12) Craig, P. N. Interdependence between Physical Parameters and Selection of Substituent Groups for Correlation Studies. *J. Med. Chem.* **1971**, *14*, 680–684.

(13) Topliss, J. G. Utilization of Operational Schemes for Analog Synthesis in Drug Design. *J. Med. Chem.* **1972**, *15*, 1006–1011.

(14) Hansch, C.; Unger, S. H.; Forsythe, A. B. Strategy in Drug Design. Cluster Analysis as an Aid in the Selection of Substituents. *J. Med. Chem.* **1973**, *16*, 1217–1222.

(15) Darvas, F. Application of the Sequential Simplex Method in Designing Drug Analogs. *J. Med. Chem.* **1974**, *17*, 799–804.

(16) Wootton, R.; Cranfield, R.; Sheppey, G. C.; Goodford, P. J. Physicochemical-Activity Relationships in Practice. 2. Rational Selection of Benzenoid Substituents. *J. Med. Chem.* **1975**, *18*, 607–613.

(17) Wootton, R. Selection of Test Series by a Modified Multidimensional Mapping Method. *J. Med. Chem.* **1983**, *26*, 275–277.

(18) Schaper, K. J. Rational Selection of Test Series for QSAR Analysis. *Quant. Struct. Act. Relat.* **1983**, *2*, 111–120.

(19) Streich, W. J.; Dove, S.; Franke, R. On the Rational Selection of Test Series. 1. Principal Component Method Combined with Multidimensional Mapping. *J. Med. Chem.* **1980**, *23*, 1452–1456.

(20) Dove, S.; Streich, W. J.; Franke, R. On the Rational Selection of Test Series. 2. Two-Dimensional Mapping of Intraclass Correlation Matrices. *J. Med. Chem.* **1980**, *23*, 1456–1459.

(21) Alunni, S.; Clementi, S.; Edlund, U.; Johnels, D.; Hellberg, S.; Sjöström, M.; Wold, S. Multivariate Data Analysis for Substituent Descriptors. *Acta Chem. Scand.* **1983**, *B37*, 47–53.

(22) van de Waterbeemd, H.; El Tayar, N.; Carrupt, P. A.; Testa, B. Pattern Recognition Study of QSAR Substituent Descriptors. *J. Comput.-Aided Mol. Des.* **1989**, *3*, 111–132.

(23) Martin, Y. C.; Panas, H. N. Mathematical Considerations in Series Design. *J. Med. Chem.* **1979**, *22*, 784–791.

(24) Wooldridge, K. R. H. A Rational Substituent Set for Structure-Activity Studies. *Eur. J. Med. Chem.* **1980**, *15*, 63–66.

(25) Austel, V. Selection of Test Compounds from a Basic Set of Chemical Structures. *Eur. J. Med. Chem.* **1982**, *17*, 339–347.

(26) de Winter, M. L. Computer Pre-selection of Compounds for Pharmacological Screening. Prediction by Fragment Description. *Eur. J. Med. Chem.* **1985**, *20*, 175–179.

(27) Boyd, D. B.; Seward, C. M. The Substituent Parameter Database: A Powerful Tool for QSAR Analysis. In *QSAR: Rational Approaches to the Design of Bioactive Compounds*; Silipo, C., Vittoria, A., Eds.; Elsevier: Amsterdam, 1991; pp 167–170.

(28) Tosato, M. L.; Geladi, P. Design: A Way to Optimize Testing Programmes for QSAR Screening of Toxic Substances. In *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W., Devillers, J., Eds.; Kluwer Academic Publishers: Dordrecht, 1990; pp 317–341.

(29) Pleiss, M. A.; Unger, S. H. The Design of Test Series and the Significance of QSAR Relationships. In *Comprehensive Medicinal Chemistry*; Ramsden, C. A., Ed.; Pergamon Press: Oxford, 1990; Vol. 4, pp 561–587.

(30) Domine, D.; Devillers, J.; Chastrette, M. A Nonlinear Map of Substituent Constants for Selecting Test Series and Deriving Structure–Activity Relationships. I. Aromatic Series. *J. Med. Chem.* **1994**, preceding paper in this issue.

(31) Sammon, J. W. A Nonlinear Mapping for Data Structure Analysis. *IEEE Trans. Comput.* **1969**, *C-18*, 401–409.

(32) Kowalski, B. R.; Bender, C. F. Pattern Recognition. A Powerful Approach to Interpreting Chemical Data. *J. Am. Chem. Soc.* **1972**, *94*, 5632–5639.

(33) Kowalski, B. R.; Bender, C. F. Pattern Recognition. II. Linear and Nonlinear Methods for Displaying Chemical Data. *J. Am. Chem. Soc.* **1973**, *95*, 686–692.

(34) Kruskal, J. B. Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrika* **1964**, *29*, 1–27.

(35) Kruskal, J. B. Comments on a "Nonlinear Mapping for Data Structure Analysis". *IEEE Trans. Comput.* **1971**, *C-20*, 1614.

(36) Wish, M.; Carroll, J. D. Multidimensional scaling and its applications. In *Handbook of Statistics*; Krishnaiah, P. R., Kanal, L. N., Eds.; North-Holland Publishing Company: Amsterdam, 1982; Vol. 2, pp 317–345.

(37) Domine, D.; Devillers, J.; Chastrette, M.; Karcher, W. Non-linear Mapping for Structure-Activity and Structure-Property Modelling. *J. Chemom.* **1993**, *7*, 227–242.

(38) Leahy, D. E.; Morris, J. J.; Taylor, P. J.; Wait, A. R. Membranes and their Models: Towards a Rational Choice of Partitioning System. In *QSAR: Rational Approaches to the Design of Bioactive Compounds*; Silipo, C., Vittoria, A., Eds.; Elsevier: Amsterdam, 1991; pp 75–82.

(39) Abraham, M. H.; Duce, P. P.; Prior, D. V.; Barratt, D. G.; Morris, J. J.; Taylor, P. J. Hydrogen Bonding. Part 9. Solute Proton Donor and Proton Acceptor Scales for Use in Drug Design. *J. Chem. Soc., Perkin Trans. 2* **1989**, 1355–1375.

(40) Thioulouse, J.; Devillers, J.; Chessel, D.; Auda, Y. Graphical Techniques for Multidimensional Data Analysis. In *Applied Multivariate Analysis in SAR and Environmental Studies*; Devillers, J., Karcher, W., Eds.; Kluwer Academic Publishers: Dordrecht, 1991; pp 153–205.

(41) Devillers, J.; Karcher, W.; Isnard, P. Graphical Display of the Fugacity Model Level I. *Chemosphere* **1991**, *23*, 633–642.
(42) Devillers, J.; Thioulouse, J.; Karcher, W. Chemometrical Evaluation of Multispecies-Multichemical Data by Means of Graphical Techniques Combined with Multivariate Analyses. *Ecotox. Environ. Safety* **1993**, *26*, 333–345.
(43) Domine, D.; Devillers, J.; Garrigues, P.; Budzinski, H.; Chastrette, M.; Karcher, W. Chemometrical Evaluation of the PAH Contamination in the Sediments of the Gulf of Lion (France). *Sci. Total Environ.* **1994**, in press.
(44) STATQSAR Package, CTIS, Lyon, France, 1993.
(45) Thioulouse, J. MacMul and GraphMu: Two Macintosh Programs for the Display and Analysis of Multivariate Data. *Comput. Geosci.* **1990**, *16*, 1235–1240.
(46) Takayama, C.; Fujinami, A.; Kirino, O.; Hisada, Y. Quantitative Structure-Activity Relationships of Antifungal 1-(3,5-Dichlorophenyl)-2,5-pyrrolidinediones and 3-(3,5-Dichlorophenyl)-2,4-oxazolidinediones. *Agric. Biol. Chem.* **1982**, *46*, 2755–2758.
(47) Li, W. Y.; Guo, Z. R.; Lien, R. J. Examination of the Interrelationship between Aliphatic Group Dipole Moment and Polar Substituent Constants. *J. Pharm. Sci.* **1984**, *73*, 553–558.