



Trade Science Inc.

# Organic CHEMISTRY

*An Indian Journal***Full Paper**

OCAIJ, 5(1), 2009 [23-30]

## Application of pattern recognition methods to determine analogy of novel molecular descriptors for drug development and synthesis

**Ronald Bartzatt**University of Nebraska, College of Arts and Sciences, Chemistry Department, Durham Science Center,  
6001 Dodge Street, Omaha, Nebraska-68182, USAE-mail : [rbartzatt@mail.unomaha.edu](mailto:rbartzatt@mail.unomaha.edu)Received: 30<sup>th</sup> December, 2008 ; Accepted: 4<sup>th</sup> January, 2009

### ABSTRACT

The design and synthesis of new pharmaceuticals is enhanced when molecular physicochemical properties are elucidated. Pattern recognition methods are ideally suited to identify the underlying relationships within a multivariate numerical matrix of molecular descriptors. Fourteen descriptors of twelve barbiturate drugs are analyzed by non-metric multidimensional scaling, discriminant analysis, cluster analysis, K-means cluster analysis, self-organizing tree algorithm (SOTA), and analysis of similarity (ANOSIM) to determine the detailed relationship of the novel descriptor LogKow(drug)/LogKow(octanol) to formula weight, polarizability, polar surface area, molecular volume, and nine other descriptors. The barbiturate class of drugs are chosen for modeling purposes due to the vast number of existing barbiturate structures and the profound variation of medicinal activity effectuated by minor structural changes. Non-metric multidimensional scaling clearly shows that the ratio LogKow(drug)/LogKow(octanol) is extremely similar to descriptors of polarizability, index of refraction, number of oxygens, nitrogens, number of hydroxyl (-OH) and amine groups (-NH<sub>n</sub>) in drug structures. Likewise hierarchical cluster analysis determined this same conclusion. In addition, the descriptor parachor is profoundly distinct from the remaining thirteen descriptors, but with formula weight and molar volume strongly similar to each other. Discriminant analysis determined that the novel descriptor LogKow(drug)/LogKow(octanol) is most dissimilar to descriptors molar refractivity, molar volume, formula weight, polar surface area, molecular volume, parachor, and molecular area. ANOSIM determined that this group of twelve barbiturates are moderately dissimilar from each other based upon the fourteen descriptors utilized for this study.

© 2009 Trade Science Inc. - INDIA

### KEYWORDS

Pattern recognition;  
Barbiturates;  
Descriptors;  
Multivariate analysis.

### INTRODUCTION

Isosterism is the relationship between sets of atoms or groups of atoms that bestow similar properties to them<sup>[1]</sup>, and the substitution of an atom or group of at-

oms is essentially an isosteric replacement. Two fundamental reasons dictate the reasoning for altering a structure are: (1) To change the characteristics of a group of atoms relative to those already present in the molecule, and (2) To influence the physical and biological proper-

## Full Paper

ties of the whole molecule<sup>[1]</sup>. Therefore the new group of atoms will have their own attributes but also induce change in molecular size, shape, and steric character of the whole molecule<sup>[1]</sup>. Minor changes in the structure can alter biological activity and these studies are referred to as structure activity relationships or SAR<sup>[2]</sup>.

Physicochemical properties (descriptors) have been defined and shown to be highly useful in predicting crucial pharmaceutical attributes. Polar surface area has been utilized successfully to predict intestinal absorption<sup>[3-5]</sup> and penetration of the blood-brain barrier (BBB)<sup>[6,7]</sup>. This is one example of the multitude of property variation induced by structure changes on the activity of a drug<sup>[2]</sup>. Multivariate statistics has been shown in previous studies to be an effective tool in predicting drug absorption based on descriptors such as polar surface area<sup>[8]</sup>, and pharmacokinetic dimensions of absorption, distribution, metabolism, and excretion<sup>[9,10]</sup>.

Molecular modeling in itself has found considerable success in predicting vitally needed anthelmintic activity for tapeworm treatment *in vivo*<sup>[11]</sup>. In addition, data mining and pattern recognition methods (i.e. hierarchical cluster analysis, multidimensional scaling, principal components analysis, and neural network modeling) have been shown to provide useful information for<sup>[12]</sup>: (a) Development of anticancer drugs; (b) Elucidation of cancer molecular pharmacology; (c) Enhancement of drug discovery for cancer care. Pattern recognition has successfully analyzed anabolic activity of steroids<sup>[13]</sup> and predicted the geno-toxicity of polycyclic aromatic compounds<sup>[14]</sup>. Specifically hierarchical cluster analysis has been found extremely useful in constructing blood-brain barrier models<sup>[15]</sup>, elucidating pharmaceutical formulation<sup>[16]</sup>, discerning pharmacokinetics<sup>[17]</sup>, analyzing microarray data<sup>[18]</sup>, and drug design<sup>[19]</sup>. Essentially this method finds hierarchy in multivariate data sets and presents results in a 2-way plotted dendrogram where subjects are grouped (clustered) together on the basis of greatest similarity<sup>[20]</sup>. Non-metric multidimensional scaling (NMMDS) is an approach for data reduction, preserves ranked differences within a multivariate data set and presents results in a 2-way plot that places subjects into groups having highest similarity<sup>[20]</sup>. K-means cluster analysis is a non-hierarchical clustering method which places subjects by greatest similarity into a predetermined number of permissible

clusters<sup>[21]</sup>. Discriminant analysis tests for separation of subjects within multivariate data to identify maximal differences among the subjects (and thereby recognize similarity among the subjects) and can also present the results in a 2-way plot<sup>[22]</sup>. ANOSIM is a non-parametric test for significant difference among groups based on distance measure<sup>[23]</sup>. A large R (up to 1) for ANOSIM indicates dissimilarity among subjects, whereas a value approaching zero suggests similarity among subjects<sup>[23]</sup>.

Barbiturates are a large group of compounds having over 2500 synthesized but with relatively few (approximately 12) being used for clinical application<sup>[24]</sup>. Their medicinal activity is used to classify members of this group by: (1) Long acting; (2) Intermediate acting; (3) Short acting; and (4) Ultra-short acting<sup>[24]</sup>. Relatively minor alterations in the structure causes substantial changes in lipid solubility and thereby medicinal activity<sup>[24]</sup>. The barbiturate drugs are an excellent choice to exhibit the profound efficacy of pattern recognition methods to determine the underlying relationships of novel descriptors to those commonly and successfully used in the important area of drug design and discovery.

## EXPERIMENTAL

### Molecular modeling methods and determination of molecular property descriptors

Two major softwares were utilized for advanced molecular modeling and included structure constituents analysis accomplished by Molinspiration (Molinspiration Cheminformatics, Liscie udolie 2, SK-841 04 Bratislava, Slovak Republic) and 2-D with 3-D analysis utilizing ChemSketch (Advanced Chemistry Development, 90 Adelaide Street West, Toronto Ontario, M5H 3V9 Canada). Some additional support was accomplished by ChemWindows 3 version 3.1.3 (Soft Shell International, 1600 Ute Ave., Grand Junction CO 81501, USA).

### Pattern recognition, multivariate analysis, and statistical analysis

Various operations to achieve pattern recognition within the numerical data matrix of molecular properties was accomplished in addition to descriptive statistics determination. Underlying relationships of the mo-

molecular properties for all compounds presented in this work were ascertained using cluster analysis and K-means cluster analysis by PAST v. 1.28 (copyright Hammer and Harper 1999-2004) and KyPlot v. 2.0 beta 15 (copyright Koichi Yoshioka 1997-2001). Data analysis of similarity, ANOSIM, was accomplished by PAST v. 1.28.

Differentiation of numerical values by group was achieved utilizing discriminant analysis and was done by KyPlot v. 2.0. Clustering using the Self Organizing Tree Algorithm or SOTA analysis<sup>[25]</sup> was accomplished by GEPAS v. 3.1 (copyright by J.Herrero, Bioinformatics Dept., 4601 Valencia, Spain). Multiple regression analysis utilizing various properties was determined by GraphPad Instat v. 3.05 (GraphPad InStat version 3.00 for Windows 95, GraphPad Software, San Diego California USA, Copyright 1992-1998 GraphPad Software www.graphpad.com).

Correlations of numerical values and descriptive statistics was done by EXCEL (Microsoft Office Excel 2003, copyright 1985-2003 Microsoft Corporation).

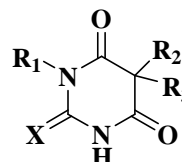
## RESULTS AND DISCUSSION

Barbiturate molecular structure contains a balance of hydrophilic and lipophilic moieties. The size and position of substituents on the pyrimidinetrione ring has substantial affect upon biological activity<sup>[24]</sup>. For this reason overseeing molecular properties is valuable for correlating physicochemical characteristics to medicinal activity. It follows that the development of novel descriptors will enhance the interface between structure and pharmaceutical activity. Lipophilicity of a drug can be effectively represented as the measure of drug distribution between n-octanol and water, or log of the quantity of drug in the organic phase relative to the aqueous phase as LogKow<sup>[2]</sup>. Because partitioning is generally based on distribution utilizing a n-octanol organic phase, the ratio of drug partitioning LogKow(drug) to the aqueous partitioning of n-octanol (or LogKow(octanol)) provides the distribution proportion of barbiturate relative to a standardized cellular membrane layer as represented by n-octanol/water by-layer. The net equation becomes LogKow(drug)/LogKow(octanol). Described differently, this ratio is the partitioning of any desired drug relative to the aqueous par-

titioning of n-octanol. A similar representation can be determined for other organic/aqueous by-layers such as water-olive oil (for blood-brain partitioning), CHCl<sub>3</sub>, 1,2-dichloroethane, and various alkanes<sup>[26]</sup>. In previous Hansch analysis the use of Log P or Log D (for ionizable drugs) was vital for prediction of biological activity<sup>[26]</sup>.

The twelve examples of barbiturate medicinals are presented in **figure 1** where the relevant substituents of the parent structure that vary are described. Four positions on the parent structure have been determined to play requisite roles in the biological activity of the barbiturate (R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub>, and X). A sulfur atom or polar oxygen atom substituted in position X are significantly correlated with medicinal action of the barbiturate<sup>[24]</sup>. In addition the carbon chains (or phenyl ring) of substituents in locations R<sub>2</sub> and R<sub>3</sub> have significant affect

Barbiturate molecular structure



	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	X
Barbital	-H	-CH <sub>2</sub> CH <sub>3</sub>	-CH <sub>2</sub> CH <sub>3</sub>	O
Hexetal	-H	-CH <sub>2</sub> CH <sub>3</sub>	-CH <sub>2</sub> -(CH <sub>2</sub> ) <sub>4</sub> -CH <sub>3</sub>	O
Amobarbital	-H	-CH <sub>2</sub> CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	-CH <sub>2</sub> CH <sub>3</sub>	O
Butabarbital	-H	-CH(CH <sub>3</sub> )CH <sub>2</sub> CH <sub>3</sub>	-CH <sub>2</sub> CH <sub>3</sub>	O
Taibutal	-H	-CH(CH <sub>3</sub> )CH <sub>2</sub> CH <sub>3</sub>	-CH <sub>2</sub> CH=CH <sub>2</sub>	O
Aprobarbital	-H	-CH(CH <sub>3</sub> ) <sub>2</sub>	-CH <sub>2</sub> CH=CH <sub>2</sub>	O
Thiopental	-H	-CH(CH <sub>3</sub> ) CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	-CH <sub>2</sub> CH <sub>3</sub>	S
Thiamyltal	-H	-CH(CH <sub>3</sub> ) CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	-CH <sub>2</sub> CH=CH <sub>2</sub>	S
Phebobarbital	-H	-CH <sub>2</sub> CH <sub>3</sub>	-Phenyl	O
Butethal	-H	-CH <sub>2</sub> CH <sub>3</sub>	-CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	O
Mephobarbital	-H	-CH <sub>3</sub>	-Phenyl	O
Methrbital	-H	-CH <sub>3</sub>	-CH <sub>2</sub> CH <sub>3</sub>	O

**Figure 1:** The diverse molecular structure of 12 medicinal barbiturates utilized in this study are shown with substituents designated for identification. Over 2500 barbiturates have been synthesized however approximately 12 satisfy current clinical requirements. The affect of substituents on medicinal activity is substantial, and provides an excellent example of the importance of size, volume, placement, and water solubility of the atoms comprising the substituent. Molecular properties are substantially affected by the choice of substituents, therefore the application of novel descriptors can be valuable for designing new clinical pharmaceuticals

## Full Paper

TABLE 1 : Molecular poperties of barbiturate drugs

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Drug	Formula weight	Molar refractivity (cm <sup>3</sup> )	Molar volume (cm <sup>3</sup> )	Parachor (cm <sup>3</sup> )	Index of refraction	Polarizability	Polar surface area(A <sup>2</sup> )	O and N	-OH and -NH	Number of rotatable bonds	Molecular volume (A <sup>3</sup> )	Molecular area (A <sup>2</sup> )	Log kow	Ratio log kow (drug)/Logkow (octanol)
Barbital	184.19	44.09	161.5	389.1	1.458	0.5	75.27	5	2	2	206.25	217.93	0.65	0.231
Hexethal	240.3	62.62	227.6	549.4	1.462	2.236	75.27	5	2	6	287.82	300.25	2.46	0.875
Amobarbital	226.27	57.95	211.4	507.3	1.46	1.743	75.27	5	2	4	265.17	275.74	2.07	0.737
Butabarbital	212.25	53.26	192.6	467.2	1.465	1.309	75.27	5	2	3	244.57	251.03	1.65	0.579
talbutal	224.26	57.62	204.2	496.4	1.476	1.525	75.27	5	2	4	259.81	267.01	1.47	0.523
Aprobarbital	210.23	52.98	187.7	456.3	1.476	1.091	75.27	5	2	3	240.27	247.41	1.15	0.409
Thiopental	242.34	65.56	207.5	548.7	1.544	1.964	58.2	4	2	4	277.52	285.45	2.85	1.01
Thiamyltal	254.35	69.96	219.1	578	1.551	2.18	58.2	4	2	5	289.7	284.16	3.23	1.15
Phebobarbital	232.24	59.21	188.1	482.8	1.541	1.401	75.27	5	2	2	254.75	262.66	1.47	0.523
Butethal	212.25	53.35	194.6	469.2	1.46	1.368	75.27	5	2	4	247.62	263.59	1.73	0.616
Mephobarbital	246.26	64.13	203.2	520.9	1.543	1.348	66.48	5	1	2	273.75	279	1.84	0.655
Methrbital	198.22	49.01	176.6	427.2	1.467	0.447	66.48	5	1	2	225.79	235.620	1.15	0.409

A = angstroms

on molecular volume, polarizability, and LogKow.

Replacement of the oxygen atom with a sulfur atom at position X results in a large decrease of total polar surface area. The total number of oxygens, nitrogens, amines, (-NH), and hydroxyl (-OH) groups regulates apparent hydrogen bond donor and acceptor properties.

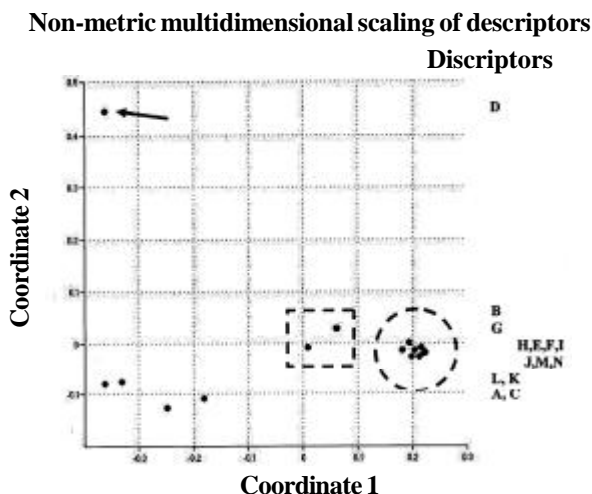
Presented in TABLE 1 are thirteen physicochemical properties of the twelve barbiturates described in figure 1 as well as the new descriptor LogKow(drug)/LogKow(octanol) designated as N, all other descriptors assigned alphabet identifiers from A to M (to be utilized in pattern recognition analysis to follow). The impact of substituent variation on numerical variance of these 14 descriptors is significant and several will be highlighted. The mean value of molar volume (C) is 197.84 cm<sup>3</sup> with standard deviation (SD) of 18.32 cm<sup>3</sup> which is 9.26% of the group mean. The mean of molecular volume (K) is 256.09 Angstroms<sup>3</sup>, with a SD of 24.98 A<sup>3</sup> which is 9.75% of the mean. Although analogous in variation, it will be shown later by pattern recognition analysis that C and K are not similar descriptors. Lipophilicity is a vital component of Hansch analysis<sup>[25]</sup>, represented as LogKow (M) has a mean of 1.81 with a SD of 0.743 which is 41.1% of the mean. Novel descriptor LogKow(drug)/LogKow(octanol) or N, has a mean of 0.6431 with a SD of 0.2641 which is 41.1% of the mean. The analogous values of SD is not unex-

pected because the numerical values of LogKow(drug) are undergoing division by a constant value of LogKow(octanol). In discussion of polar surface area (G) it is striking how a change of oxygen to sulfur in the X position (see figure 1) induces a large change in overall molecular polar surface area from group mean of 73.51 A<sup>2</sup> and SD of 3.71 A<sup>2</sup> (minus thiopental and thiamyltal) to overall group mean 70.96 A<sup>2</sup> and SD of 6.838 A<sup>2</sup>. Clearly the variation of a judiciously chosen atom can bring about significant alteration of physicochemical properties and biological activity (thiopental and thiamyltal are ultra-short acting barbiturates). Placement of aliphatic or nonaliphatic substituents on the parent structure of barbiturates (see figure 1) results in a striking variation of molecular polarizability (F) (see TABLE 1). The overall group mean of F is 1.426 cm<sup>3</sup> with a SD of 0.5702 cm<sup>3</sup> which is 40.0% of the mean. Index of refraction (E) remains highly consistent with a mean of 1.492 and SD of 0.03944 which is 2.64% of the group mean value. Parachor (descriptor D) values have a mean of 491.64 cm<sup>3</sup>, SD of 54.15 cm<sup>3</sup> which is 11.0% of the mean. With analogous variation in numerical values the group mean of molar refractivity (descriptor B) is 57.48 cm<sup>3</sup>, SD of 7.39 cm<sup>3</sup> which is 12.9% of the mean. Finally, the group mean of formula weight (descriptor A) is 223.60 with a SD of 20.90 that is 9.35% of the mean. Clearly the variation of substituents within barbiturates results in substantial variation of descriptor

numerical values and consequently their biological activity. It follows that using physicochemical properties are an important component for elucidating their medicinal activity. The application of new descriptors can improve understanding of the association of drug design and biological activity. Grubb's test for outliers among the numerical values of novel descriptor N ( $\text{LogKow}(\text{drug})/\text{LogKow}(\text{octanol})$ ) indicated no outliers among the values calculated for these 12 barbiturates.

While preserving the ranked differences among the subjects of a multivariate matrix non-metric multidimensional scaling (NMMDS) groups (clusters) the subjects with others that have the greatest similarity. Distances between subjects projected into a 2-way plot indicates relative dissimilarity (ie. Closest subjects have higher similarity than those far apart). The results of NMMDS presented in figure 2 represent analysis of descriptors A to N and not the barbiturates as subjects. At great distance from all the remaining descriptors, parachor (D) (see inset arrow) is determined to be highly dissimilar from thirteen other descriptors. The numerical values of parachor (mean =  $491.64 \text{ cm}^3$ ) are considerably larger than all the remaining descriptors. Parachor is a measure of molecular polarizability and van-der-Waals interactions<sup>[2]</sup>. Notably the descriptors molar refractivity (B) and polar surface area (G) (see inset rectangle of Figure 2) are similar to each other but clearly distinct from the supercluster (see inset circle) containing H, E, F, I, J, M, and N. Novel descriptor N ( $\text{LogKow}(\text{drug})/\text{LogKow}(\text{octanol})$ ) is determined to have greatest similarity to properties of hydrogen bond interaction (H, I) and properties of topology (J) but considerably different from properties representing polarizability: D (parachor), B (molar refractivity), and C (molar volume). Other descriptors dissimilar to N are molecular area (L) and molecular volume (K). These results indicate that N would be useful in providing additional information of a drug that balances contributions of molar refractivity (B), molecular area (L), molar volume (C), and molecular volume (K).

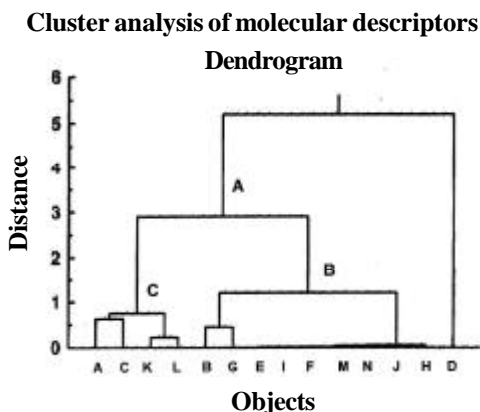
The hierarchical cluster analysis results are presented in the form of a dendrogram, see figure 3, utilizing divisive clustering with standard Euclidean distance (ie. Distance between subjects is the smallest value to connect them). Here single linkage clustering was applied, or computation of similarity by the closest pairs of ob-



**Figure 2:** Non-metric multidimensional scaling results are presented in a 2-way plot to project the distance differences among subjects of a multivariate matrix. Clearly shown here are the relative similarities among the descriptors presented in TABLE 1 and defined alphabetically A through N (see TABLE 1). Interestingly, parachor (D) is determined to be highly dissimilar from all other descriptors (see inset arrow). Molar refractivity (B) and polar surface area (G) are most similar to each other (see inset rectangle). The novel descriptor  $\text{LogKow}(\text{drug})/\text{LogKow}(\text{octanol})$  or N is found among a cluster of descriptors (inset circle) including  $\text{LogKow}(\text{drug})$  (M), number of rotatable bonds (J), number of -OH and -NH (I), polarizability (F), index of refraction (E), and number of oxygens and nitrogens (H)

servations between two groups. Objects falling within the same cluster are determined to be most similar. Again, as in figure 2, descriptor D (parachor) is shown to be highly distinct for other descriptors and falls within a cluster of its own. A supercluster is divided at node A into two large clusters that are further divided at node B and node C. For objects are categorized under node C into two clusters having two objects each. Descriptors A (formula weight) and C (molar volume) are most similar, however descriptors K (molecular volume) and L (molecular area) are paired together and thereby most similar. Thus far these results are analogous to those of non-metric multidimensional scaling (see Figure 2). The remaining nine descriptors fall under node B. However further hierarchical division places molar refractivity (B) and polar surface area (G) into a distinct cluster from the remaining seven. Finally, descriptors E, I, F, M, M, and H are shown to be most similar to the novel descriptor  $\text{LogKow}(\text{drug})/\text{LogKow}(\text{octanol})$  (or N). So

## Full Paper



**Figure 3:** The dendrogram presented shows results of cluster analysis of the data matrix that is TABLE 1 and utilizing single linkage clustering with standard Euclidean distance. Analogous to non-metric multidimensional scaling it shows parachor (D) to be highly distinct from all other descriptors. Again formula weight (A) and molar volume (C) are determined to have highest similarity as is molecular volume (K) and molecular area (L). The novel descriptor LogKow(drug)/LogKow(octanol) (N) is grouped by high similarity to index of refraction (E), -OH and -NH (I), polarizability (F), LogKow(drug) (M), number of rotatable bonds (J), and oxygens and nitrogens (H)

descriptors A, C, K, L, D, B, and G are dissimilar from N, therefore the parameter LogKow(drug)/LogKow(octanol) is not ambiguous of those molecular features comprising A, C, K, L, D, B, and G.

ANOSIM one-way determination preserves the differences among groups of a multi-variate data by converting distances to ranks<sup>[23]</sup>. An R value of 0.3333 is acquired for the numerical values of 14 descriptors of TABLE 1, which indicates a lower level of dissimilarity overall. This result is reasonable for a measurement from a population of drugs from an identical class. However the affect on structure properties from substituent variation is significant from drug to drug, this is clearly observed by investigating the numerical differences among drug descriptors. The underlying relationships are made clear through analysis that utilize pattern recognition methods.

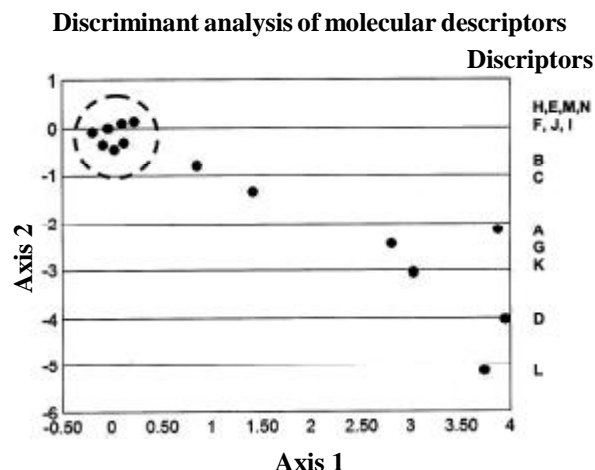
The strength of linear relationship among variables is often determined by applying Pearson's correlation analysis that produces the coefficient  $r$ , where zero indicates no correlation, negative values indicate inverse correlation, and  $r > 0.9000$  indicates a very high correlation. A high correlation does not prove causality, but describes a analogous trend in direction (ie. As one

variable increases another also increases).

For descriptors presented in TABLE 1 a high correlation ( $r > 0.7000$ ) was determined for formula weight (A), molar refractivity (B), molar volume (C), parachor (D), index of refraction (E), polarizability (F), molecular volume (K), molecular area (L), LogKow (M), and LogKow(drug)/LogKow(octanol) (N). The determination of LogKow incorporates the molecular size while summing contributions to aqueous/organic by-layer partitioning from component substituents. Interestingly the novel descriptor N retains the high correlation to formula weight, molar refractivity, etc; and therefore N will retain some facet of molecular size influence onto the biological activity. No correlation exists between N and I, with an inverse correlation between N and descriptors G and H.

K-means cluster analysis is a non-hierarchical cluster method in which subjects are clustered according to similarity (analogous to hierarchical cluster analysis, see figure 3) but the number of clusters are specified by the user<sup>[21]</sup>. Because K-means clustering does not require prior computation of a proximity of distance/similarity in every case this method can analyze much larger data sets than for hierarchical clustering analysis.

Results of K-means clustering of TABLE 1 reveals that molar refractivity (B) is most similar to polar surface area (G). Again parachor (D) is distinct from the remaining 13 descriptors as indicated in hierarchical cluster analysis and non-metric multidimensional scaling. The number of oxygens plus nitrogens (H) is most similar to number of amines and hydroxyls (I). The remaining descriptors are divided among two super clusters; the first grouping of formula weight (A), molar volume (C), molecular volume (K), and molecular area (L); and inclusive of LogKow(drug)/LogKow(octanol) (N), LogKow (M), number of rotatable bonds (J), polarizability (F), and index of refraction (E). Likewise self organizing tree analysis<sup>[25]</sup>(SOTA) placed the following descriptors into grouping indicating highest numerical similarity: LogKow(drug)/LogKow(octanol) (N), LogKow (M), number of rotatable bonds (J), polarizability (F), and index of refraction (E). The remaining nine descriptor were placed individually into separate clusters. The results of SOTA analysis are heavily influenced by numerical congruency seen with the supercluster of descriptors that contain N and hav-



**Figure 4:** Discriminant analysis is highly sensitive to distance measurements of objects within a multivariate data matrix. The novel descriptor LogKow(drug)/LogKow(octanol) (N) is contained in a super cluster of descriptors (see inset circle) determined to have the greatest similarity and includes index of refraction (E), LogKow(drug) (M), oxygens and nitrogens (H), -OH and -NH (I), number of rotatable bonds (J), and polarizability (F). All other descriptors are determined to have substantial dissimilarity and are strewn over the 2-way with substantial distance between each other and the super cluster

ing numerical values less than ten.

SOTA analysis has substantial determinateness but reveals formally the expected outcome accompanying standard descriptors applied in drug design. Multiple regression analysis was performed utilizing formula weight (A) as the dependent variable and molar refractivity (B), molar volume (C), parachor (D), and novel descriptor LogKow(drug)/LogKow(octanol) (N) as independent variables. A multiple regression equation meliorates the design of new drugs by determining relationships among molecular properties that are vital for biological activity. The resulting equation appears as follows:  $A = 56.579 + [B][5.361] + [C][0.7074] - [D][0.5746] + [N][1.780]$ . The  $R^2$  for this equation is 0.9962, accounting for 99.62% of the variance in formula weight. A direct Pearson correlation between (N) and formula weight is 0.8580 (considered high correlation), where a straight linear fit becomes  $y = 0.01085x - 1.78$ . In addition, the multiple regression equation defines numerically the relationship of formula weight of barbiturate drugs to polarizability, van-der-Waals interactions, and topology of molecular structure.

The dynamic results of discriminant analysis per-

mits the confirmation or rejection of hypothesis that two subjects are distinct (ie. Dissimilar)<sup>[22]</sup>. A 2-way plot of scores along axis 1 and axis 2 will present the subjects with Euclidean distance between them as approximations of dissimilarity. A 2-way plot of scores for all descriptors of TABLE 1 is presented in figure 4 along axis 1 and 2. Descriptor N is found in close proximity (and therefore similar) to index of refraction (E), LogKow (M), oxygens and nitrogens (H), polarizability (F), number of rotatable bonds (J), and number of hydroxyls and amines (I) (see inset circle). All remaining descriptors are spread out over the range of axis 1 and 2 and having greater distance between them (higher distinction). Clearly descriptor N is not similar to B, C, A, G, K, D, and L.

## CONCLUSION

A diverse but clinically relevant population of barbiturate drugs were utilized to demonstrate the underlying relationships of 14 descriptors including a novel descriptor (N), a ratio that compares standard drug partitioning LogKow to actual partitioning of n-octanol (which itself is generally taken to represent lipid by-layer partitioning). Various pattern recognition techniques were applied to clarify underlying relationships found within the data matrix of 14 descriptors (TABLE 1). Non-metric multidimensional scaling analysis determined that descriptor N (LogKow(drug)/LogKow(octanol)) is highly similar to those found in supercluster (see inset circle of Figure 2) containing H, E, F, I, J, and M. Likewise to NMMDS, hierarchical cluster analysis (Figure 3) associated N to the same descriptors. Even discriminant analysis determined that N has negligible distinction (ie. Higher similarity) from descriptors H, E, F, I, J, and M (see Figure 4). The use of pattern recognition methods showed conclusively that novel descriptors can be related to other diverse and disparate physicochemical properties to find underlying relationships, similarities, and consequently-dissimilarities. The novel descriptor N is a numerical ratio between standard organic-aqueous partitioning LogKow and the partitioning of n-octanol-which is accepted to represent a lipid by-layer model. The final result is a numerical value that describes a drug's partitioning to the by-layer model. Other models can be utilized, such as water-olive oil (for blood-

## Full Paper

brain partitioning),  $\text{CHCl}_3$ , 1,2-dichloroethane, and various alkanes<sup>[26]</sup>. Grubb's test showed no outliers exist among values of N. Pearson's correlation coefficient outcome showed that N is highly correlated to the numerical values of all descriptors except with I (no correlation) and with an inverse correlation between N and descriptors G and H. ANOSIM identified a moderate level of dissimilarity among all the descriptors of TABLE 1, an outcome that reflects the effects of substituent variation among an otherwise indistinguishable group of barbiturate drugs. The ratiocination of this study is that pattern recognition methods are highly effective in determining relationships between novel descriptors and the large body of physicochemical properties applied to drug design and development. In addition, the ratio  $\text{LogKow}(\text{drug})/\text{LogKow}(\text{octanol})$  is related to properties than assort polarizability and van-der-Waals molecular traits. This ratio adverts drug partitioning relative to the generally applied membrane by-layer modeling accomplished with the n-octanol/water system.

### ACKNOWLEDGMENTS

This work was supported by the College of Arts and Sciences, University of Nebraska, Department of Chemistry, Durham Science Center, Omaha NE 68182 USA.

### REFERENCES

- [1] L.B.Klier, L.H.Hall; *Chemistry and Biodiversity*, **1**, 138 (2004).
- [2] G.Thomas; *Medicinal Chemistry*, John Wiley and Sons; New York, (2000).
- [3] K.Palm, P.Stenberg, K.Luthman, P.Artursson; *Pharmaceutical Research*, **14**(5), 568 (1997).
- [4] S.Winiwarter, N.M.Bonham, F.Ax, A.Hallbeg, H.Lennemas, A.Karlen; *J.Med.Chem.*, **41**, 4939 (1998).
- [5] D.E.Clark; *J.Pharm.Sci.*, **88**, 807 (1999).
- [6] D.E.Clark; *J.Pharm.Sci.*, **88**, 815 (1999).
- [7] J.Kelder, P.D.J.Grootenhuys, D.M.Bayada, L.P.C.Delbressine, J.P.Ploemen; *Pharm.Res.*, **16**, 1514 (1999).
- [8] W.J.Egan, K.M.Merz, J.J.Baldwin; *J.Med.Chem.*, **43**(21), 3867 (2000).
- [9] J.F.Blake; *Current Opinion in Biotechnology*, **11**, 104 (2000).
- [10] V.Daralis, A.Tsantill-Kakoulidou, P.Macheras; *Pharmaceutical Research*, **19**(12), 1827 (2002).
- [11] K.B.Lipkowitz, R.O.McCracken; *Parasitology Research*, **79**(6), 475 (1993).
- [12] L.M.Shi, Y.Fan, J.K.Lee, M.Waltham, D.T.Andrews, U.Scherf, K.D.Paull, J.N.Weinstein; *J.Chem.Inf.Comput.Sci.*, **40**(2), 367 (2000).
- [13] Y.Alvarez-Ginarte, Y.Marrero-Ponce, J.Ruiz-Garcia, L.Montero-Cabrera, J.Garcia de la Vega, P.Noheda, R.Crespo-Otero, F.Zaragoza, R.Garcia-Domenech; *J.Comput.Chem.*, **29**(3), 317 (2008).
- [14] L.He, P.Jurs, L.Custer, S.Durham, G.Pearl; *Chemical Research in Toxicology*, **16**(12), 1587 (2003).
- [15] D.Pan, M.Lyer, J.Liu, Y.Li, A.Hopfinger; *Journal of Chemical Information and Computer Sciences*, **44**(6), 2083 (2004).
- [16] S.Horhota, C.Aitken; *Journal of Pharmaceutical Sciences*, **80**(1), 85 (1991).
- [17] M.Fossier, K.Chang, D.Young; *Acta Pharmaceutica Jugoslavica*, **402**(2, Supp 1), 225 (1990).
- [18] W.Shannon, R.Culverhouse, J.Duncan; *Pharmacogenomics*, **4**(2), 229 (2003).
- [19] M.Moreno, N.Magalhaes, C.Cavalcanti, A.Alves; *Quimica Nova*, **19**(6), 594 (1996).
- [20] D.A.T.Harper; *Numerical Palaeobiology*, John Wiley and Sons, New York, (1999).
- [21] S.T.Bow; *Pattern Recognition*, Marcel Dekker; New York, (1984).
- [22] J.C.Davis; *Statistics and Data Analysis in Geology*, John Wiley and Sons, New York, (1986).
- [23] K.R.Clarke; *Australian Journal of Ecology*, **18**, 117 (1993).
- [24] L.Goodman, A.Gilman; 'The Pharmacological Basis of Therapeutics', 4<sup>th</sup> Edition, The MacMillan Company; London, (1970).
- [25] J.Herrero, A.Valencia, J.Dopazo; *Bioinformatics*, **17**, 126 (2001).
- [26] A.Avdeef; *Absorption and Drug Development*, Wiley-Interscience, Hoboken, (2003).