Applying Pattern Recognition Methods to Analyze the Molecular Properties of a Homologous Series of Nitrogen Mustard Agents

Submitted: July 6, 2005; Accepted: February 6, 2006; Published: April 14, 2006

Ronald Bartzatt[1] and Laura Donigan[1]

[1]University of Nebraska, Durham Science Center, Department of Chemistry, Laboratory of Pharmaceutical Studies, Omaha, NE

ABSTRACT

The purpose of this research was to analyze the pharmacological properties of a homologous series of nitrogen mustard (N-mustard) agents formed after inserting 1 to 9 methylene groups (-CH_2-) between 2 -$N(CH_2CH_2Cl)_2$ groups. These compounds were shown to have significant correlations and associations in their properties after analysis by pattern recognition methods including hierarchical classification, cluster analysis, nonmetric multi-dimensional scaling (MDS), detrended correspondence analysis, K-means cluster analysis, discriminant analysis, and self-organizing tree algorithm (SOTA) analysis. Detrended correspondence analysis showed a linear-like association of the 9 homologs, and hierarchical classification showed that each homolog had great similarity to at least one other member of the series—as did cluster analysis using paired-group distance measure. Nonmetric multi-dimensional scaling was able to discriminate homologs 2 and 3 (by number of methylene groups) from homologs 4, 5, and 6 as a group, and from homologs 7, 8, and 9 as a group. Discriminant analysis, K-means cluster analysis, and hierarchical classification distinguished the high molecular weight homologs from low molecular weight homologs. As the number of methylene groups increased the aqueous solubility decreased, dermal permeation coefficient increased, Log P increased, molar volume increased, parachor increased, and index of refraction decreased. Application of pattern recognition methods discerned useful interrelationships within the homologous series that will determine specific and beneficial clinical applications for each homolog and methods of administration.

KEYWORDS: antineoplastic, nitrogen mustards, multivariate, pattern recognition.

INTRODUCTION

Statistical analysis methods are applied to determine the source of property differences, side effects, and variation in activities for drugs that have similar structural features.[1] One important approach to these studies is the alteration of the size and shape of a drug by instituting the following: (1) changing the number of methylene groups; (2) changing the degree of unsaturation; and (3) adding or removing a ring system.[1] The addition of methylene groups along a chain increases the lipophilicity of a drug, increases permeation of lipid cell membranes, and may increase the activity.[1] The addition of a methylene group (-CH_2-) along a chain constitutes the formation of a homologous series.[2] Many studies have shown that the addition of 1 to 6 or 7 methylene groups results in the increase of drug activity[2]; however, the chain lengthening beyond this point results in a decrease of activity. A 2-way plot of the number of methylene groups versus the drug potency results in a unimodal figure.[2]

Statistical analysis of properties can elucidate beneficial characteristics of a drug candidate before they are put through expensive and time-consuming biological testing.

Covalent DNA binding drugs include 6 groups of therapeutics: (1) nitrogen mustards (N-mustards), (2) aziridines, (3) alkane sulfonates, (4) nitrosoureas, (5) platinum compounds, and (6) methylating agents.[3] N-mustards are a group of highly reactive compounds that form covalent bonds with certain heteroatomic nucleophilic groups that can bear nitrogen, oxygen, or sulfur atoms.[2] These alkylating compounds are considered cell-cycle nonspecific.[2] The relative rate of nucleophilic substitution-type alkylation reaction is of the general order thiolate > amino > phosphate > carboxylate.[2] The preferred nucleophilic sites of DNA are N - 7 of guanine > N - 3 of adenine > N - 1 of adenine > N - 1 of cytosine.[2] Nucleophilic sites on RNA, protein, and other biomolecules may be alkylated by N-mustards,[3] which is also considered cytotoxic action and may be responsible for undesired side effects.[1] Previous studies have shown that many aromatic N-mustard agents will alkylate by an SN1 mechanism[4] (unimolecular nucleophilic substitution) and aliphatic N-mustards form an ethylene immonium ion intermediate.[4] Differences in lipophilicity, solubility, medicinal activity, and drug side effects is considered to be dependent on the physiochemical features.[5] Structural modifications of an alkylating agent can allow the drug to be administered orally as well as parenterally.[5] A goal of this study was to elucidate favorable structural features within

Corresponding Author: Ronald Bartzatt, University of Nebraska, Durham Science Center, Department of Chemistry, Laboratory of Pharmaceutical Studies, 6001 Dodge St, Omaha, NE 68182. Tel: (402) 554-3612; Fax: (402) 554-3888; E-mail: bartzatt@mail.unomaha.edu

this group of anticancer agents by using pattern recognition methods.

For this study of molecular properties, 2 groups of statistical tools were used: (1) pattern recognition,[5] and (2) correlation analysis.[5] Previous studies have demonstrated versatile designs for N-mustard agents based on naphthalimide,[6,7] (L)-Carnitine,[8] steroid derivatives,[9,10] tallimustine,[11] tetrapyrrole,[12] aromatic structures[13] (including nitroaniline,[14] aminocinnamic acid[15]), imidazole derivatives,[16] and distamycin.[17] Alkylating agents are the largest class of anticancer agents and are cell-cycle nonspecific drugs that form highly reactive electrophilic species. Alkylating agents may be monofunctional (1 reactive group) or bifunctional (2 reactive groups), leading to monoalkylation and single-strand breaks in DNA or cross-linking, respectively. Each of the drug designs above showed variation in properties owing to molecular structure, which were beneficial for their specific clinical application. The analysis of physiochemical properties has been shown to reveal the effects of basicity and structural substituents.[18] Two of the homologous agents studied in this work have been produced previously,[19] with other members following similar synthetic approach. Pattern recognition methods will show important and beneficial pharmacological interrelationships, which support the assertion that these homologs have potential for clinical application.

Materials and Methods

Statistical Algorithms

Descriptive statistics and correlation establishes numerical relationships, which help discern pharmacological characteristics. Software used included Quattro Pro 12 (Corel Corporation, copyright 1996-2004, Ottawa, Ontario, Canada) and KyPlot Version 2.0 beta 15 (copyright 1997-2001, Koichi Yoshioka).

Determination of Molecular Modeling and Molecular Properties

Molecular modeling elucidates numerical values of structural characteristics including size, polarizability, and hydrophobicity. Molecular modeling and determination of molecular properties of drug structures was accomplished by ChemSketch (Advanced Chemistry Development, Toronto, Ontario, Canada), Molinspiration (Molinspiration, Bratislava, Slovak Republic), Actelion (Actelion, Allschwil, Switzerland), and MolSoft (MolSoft, La Jolla, CA). Solubility, Log Kow, and dermal permeation coefficient were determined by EpiSuite software (AllidSystems, Sylmar, CA). Drug likeness was determined by methods of Actelion and MolSoft. Values of pK_a were determined by using SPARC On Line Calculator for properties (Version August 2003, University of Georgia, Athens, GA, www.uga.edu).

Pattern Recognition Analysis and Multiple Regression Algorithms

Multivariate data matrices were analyzed by these algorithms to show pattern relationships within the numerical values. Multiple regression analysis was determined by GraphPad InStat Version 3.0 (Graph-Pad Software, San Diego, CA) and Smith's Statistical Package Version 2.5 (Copyright 1995-2001, Gary Smith, Pomona College, Clairemont, CA). SOTA analysis accomplished by GEPAS Version 1.1 (Department of Bioinformatics, Centro de Investigacion Principe Felipe, Valencia, Spain).[20] Discriminant analysis, nonmetric MDS, detrended correspondence analysis, and K-means cluster analysis determined by PAST Version 0.45 (copyright May 2001, Oyvind Hammer, D. A. T. Harper). Cluster analysis was determined by KyPlot. Hierarchical classification was accomplished by StatBox Version 2.5 (Grimmer Logiciels, Neuilly-Sur-Seine, France). Analysis of similarities (ANOSIM) for properties of the homologous series was accomplished by PAST Version 045.

Results and Discussion

The N-mustard agents studied here are identical twin drugs that have 4 sites of potential covalent alkylation (ie, 2 bifunctional constituents). Two members of the homologous series have been synthesized, and the reaction mechanism has been determined by using 2 homologs, which have 2 or 6 methylene groups within the aliphatic carbon chain[19] (see Figure 1). Analysis of the molecular properties showed significant levels of correlation and association. These observations led to the determination of a homologous series in which additional methylene groups between the N-mustard groups contribute to beneficial pharmacological parameters.[1] Determination of their molecular properties (eg, molar volume, parachor, Log P, polar surface area [PSA]) indicated linear correlations and similarities within the series, suggesting significant potential as clinical therapeutics.[19] A homologous series of drugs show enhanced pharmacological activity up to the addition of 6 or 7 methylene groups; however, the benefits then decrease after greater than 7 or 8 methylene groups. Structure-property correlations associate any molecular property to any other property and facilitate the design or improvement of clinical drugs.

The general molecular structure of the homologs is presented in Figure 1. The major constituents of the agents are 2 bifunctional N-mustard groups connected by an aliphatic carbon chain composed of 1 to 9 methylene groups. There are a total of 4 sites of possible covalent alkylation activity per molecule. The length of the aliphatic carbon chain significantly influences their molecular properties.

Numerical values of important pharmacological properties are shown in Table 1 for each homolog. They are identified

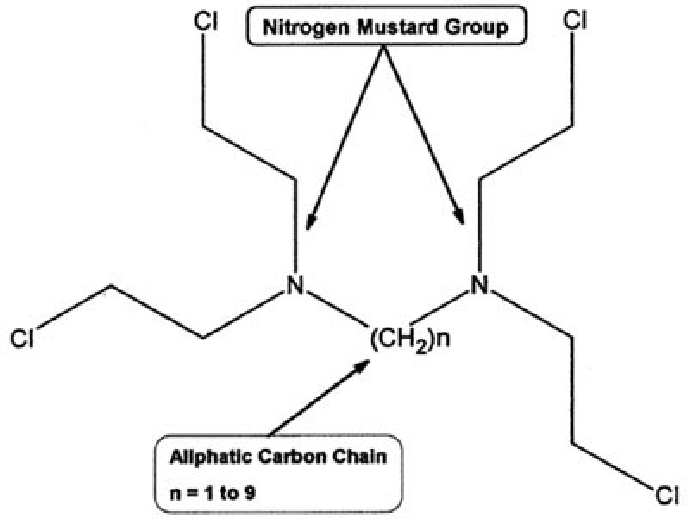

Figure 1. Molecular structure of N-mustard homolog series is shown here with 2 end N-mustard groups. For this study the number of methylene groups varies from 1 to 9.

by the number of methylene groups (-CH$_2$-) in the aliphatic chain (seen on left-hand side). All properties shown in Table 1 vary according (high correlation) to the number of methylene groups present in homologs. The Pearson correlation coefficient (*r*) for formula weight to each of molar volume, molar refractivity, parachor, and number of rotatable bonds is 1.000. Correlation (*r*) of formula weight to index of refraction for all homologs is between −0.9800 and −1.000. Value of PSA remains constant at 6.476 A^2 and is the summation of area for nitrogen and oxygen atoms with their hydrogens attached.[21] These PSA numerical values indicate that all homologs are expected to be more than 95% absorbed from the intestinal tract.[21,22] In addition, the PSA values indicate all 9 series members will readily cross the blood-brain barrier[23] (PSA values are considerably less than 90 A^2). Molar volume, molar refractivity, and parachor are polarizability parameters. The number of rotatable bonds is a simple topological parameter, which is a measure of molecular flexibility that increases as the number of methylene groups increases. Index of refraction is the speed of light in a vacuum divided by the speed of light within the agent, these values having a range of 0.013. Molar refractivity is a measure of steric factors, a constitutive-additive property, and a measure of the volume occupied by a group of atoms. Molar refractivity increases as the formula weight and molar volume increases, indicating the concurrent increase in steric effects.

Numerical values of partition coefficients are determined by various methods for each homolog and presented in Table 2 (homologs are identified by number of methylene groups within the aliphatic carbon chain, seen on left-hand side). C Log P is determined by summation of contributions from molecular constituents. Values of Log Kow are obtained presuming all species present in the aqueous and organic layers are neutral. Partition of drugs into the central nervous system is indicated by values of Log BB, (BB indicates Cbrain/Cblood) and is calculated from the following relationship[23]:

$$Log\ BB = 0.0148\ (PSA) + 0.152\ (Clog\ P) + 0.139 \qquad (1)$$

Values of the partition coefficients are highly correlated to the number of methylene groups within the homologs. Log BB values for all homologs are greater than 0.3, which indicates all 9 homologs will readily cross the blood-brain barrier to attack brain tumors.[23] Determining the numerical values of such critical parameters reduces the time and expense required for drug design.[23] The application of pattern recognition analysis will also illuminate beneficial dissimilarities within this group of compounds. The means for calculated values of miLog P, Clog P, and Log Kow for each homolog as the number of methylene groups increases from 1 to 9 are 2.36, 2.47, 2.94, 3.40, 3.86, 4.33, 4.79, 5.25, and 5.71, respectively. The Pearson correlation coefficient for all partitioning coefficients to the number of methylene groups is greater than 0.9800 for all homologs.

Table 1. Molecular Properties of Homologous Series*

Number of Methylene Groups in Aliphatic Chain	Formula Weight	Molar Volume	Molar Refractivity	Parachor	Index of Refraction	Polar Surface Area	Number of Rotatable Bonds
1	296.1	237.0	70.58	595.3	1.507	6.476	10
2	310.1	253.6	75.21	635.1	1.504	6.476	11
3	324.1	270.1	79.84	674.9	1.502	6.476	12
4	338.1	286.6	84.48	714.7	1.501	6.476	13
5	352.2	303.1	89.11	754.4	1.499	6.476	14
6	366.2	319.6	93.74	794.2	1.498	6.476	15
7	380.2	336.1	98.38	834.0	1.497	6.476	16
8	394.3	352.6	103.0	873.8	1.495	6.476	17
9	408.3	369.1	107.6	913.6	1.494	6.476	18

*Units for molar volume, molar refractivity, and parachor are cm^3. Units for polar surface area are Angstroms.^{2}
Nomenclature for homologs: homolog 1 is N,N,N,N-tetrakis(2-chloroethyl)methanediamine; and so forth to homolog 9 N,N,N, N-tetrakis(2-chloroethyl)nonane-1,9-diamine.

Table 2. Partition Coefficients of Homologous Series*

Number of Methylene Groups in Aliphatic Chain	miLog P	Clog P	Log Kow	Log BB
1	2.64	2.22	2.23	0.381
2	2.37	2.34	2.72	0.399
3	2.80	2.81	3.21	0.470
4	3.23	3.27	3.70	0.540
5	3.67	3.73	4.19	0.610
6	4.10	4.20	4.68	0.682
7	4.54	4.66	5.17	0.752
8	4.97	5.13	5.66	0.823
9	5.40	5.59	6.15	0.893

*Values of milog P calculated by Molinspiration cheminformatics. Clog P calculated by Actelion cheminformatics. Log Kow calculated by EpiSuite. Log BB calculated from Log BB = −0.0148(PSA) + 0.152(Clog P) + 0.139, where PSA indicates polar surface area.

Previous studies have shown that drugs having Log P values at 2 ± 0.5 will readily cross the blood-brain barrier.[24] Therefore, homologs having 1 and 2 methylene groups are shown to have potential as therapeutics for treatment of brain tumors. In addition, by Log P value, the homologs 1, 2, and 3 are suitable for oral formulation and dosing. Likewise, homologs 4 and 5 are suitable for transdermal administration ($3 < \text{Log P} < 4$). Homolog 9 would be suitable for sublingual absorption. Lipophilicity as indicated by octanol/water partition coefficient Log P conveys information of hydrophobic interactions or the tendency of a molecule to be solvated by water. Any ionization of a drug in vivo influences cell membrane permeation (ie, ionized drugs do not readily penetrate lipid bilayers) and can seriously affect the medicinal activity of a drug. The pK_a's of all homologs were calculated and fell in a range from 4.06 to 6.08 from homologs 1 to 9, respectively (considering each homolog to be a base and ionization occurring at the N atom only). Using these values and the normal pH of blood at 7.4, the percentage ionization of all homologs remains below 15% at pH 7.4. Ionization above 90% appears in the pH range of 1 to 5. Zero percent ionization occurs for all homologs at pH greater than 8. Significant ionization (greater than 50%) of total amount of drug present occurs only at a pH range from pH 4 (for homolog 1) to pH 6 (for homolog 9). These findings further demonstrate the significant potential and anticipated effectiveness of these homolog N-mustard agents to penetrate lipid bilayers and express anticancer activity.

SOTA analysis will assemble subjects (ie, drugs) into clusters having highest similarity, and in this respect is analogous to cluster analysis.[20] SOTA analysis performed on properties shown in Table 1 results in 2 clusters bearing the most similar homologs (identification by number of methylene groups in aliphatic chain): Cluster 1) 1, 2, 3, 4; Cluster 2) 5, 6, 7, 8, 9. Homologs 1 to 4 are seen as most similar and distinct from homologs 5 to 9. K-means cluster analysis will likewise designate subjects into clusters of highest similarity but limits the number of clusters to a target value designated by the investigator (there is no hierarchy). In this study, K-means cluster analysis was performed on properties shown in Table 1 into a maximum of 3 clusters. Results of K-means cluster analysis are as follows (homologs identified by number of methylene groups): Cluster 1) 1, 2, 3; Cluster 2) 7, 8, 9; and Cluster 3) 4, 5, 6. Given 2 sets of multivariate data matrices, then discriminant analysis will select out the subjects in a manner that will maximize the differences among them.[25] Discriminant analysis performed on properties listed in Table 1 formulated the following 2 groups of homologs (identification by number of methylene groups in aliphatic chain): Group 1) 1, 2, 3, 4, 5; and Group 2) 6, 7, 8, 9. Therefore homologs having higher molecular weight (6, 7, 8, and 9) are distinguishable from homolog members having smaller molecular weight (1, 2, 3, 4, and 5). This finding is corroborated by the results of analysis of similarities (ANOSIM) of Table 1. ANOSIM provides a statistical analysis to determine whether a significant difference exists between 2 or more groups. A positive test statistic *R* value approaching 1 indicates dissimilarity between groups. An *R* value of 0.7219 was obtained between the higher molecular weight homologs 6, 7, 8, and 9 when compared with lower molecular weight homologs 1, 2, 3, 4, and 5, thus indicating significant dissimilarity.

Hierarchical classification constructs a tree of classifications, which is modeled through creating associations between subordinate and superordinate classes.[26] Subjects are placed within the most appropriate class and it is a highly sensitive analysis method. Figure 2 presents results of

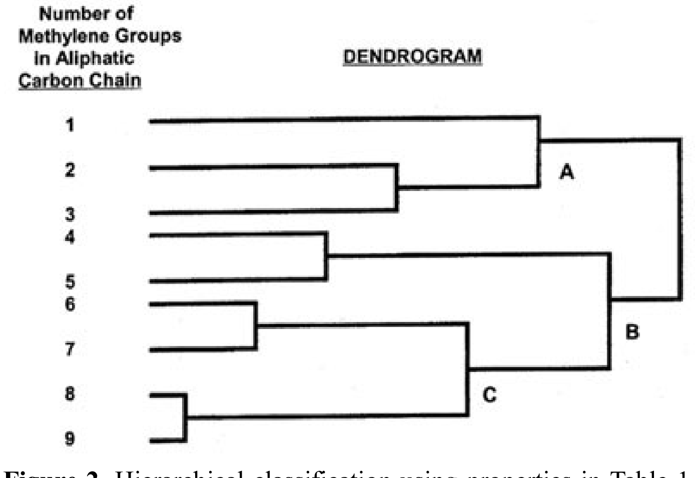

Figure 2. Hierarchical classification using properties in Table 1. Homologs having 1, 2, and 3 methylene groups fall under node A and homologs having 4 to 9 fall under node B. Similar homologs are as follows: 2 and 3, 4 and 5, 6 and 7, 8 and 9.

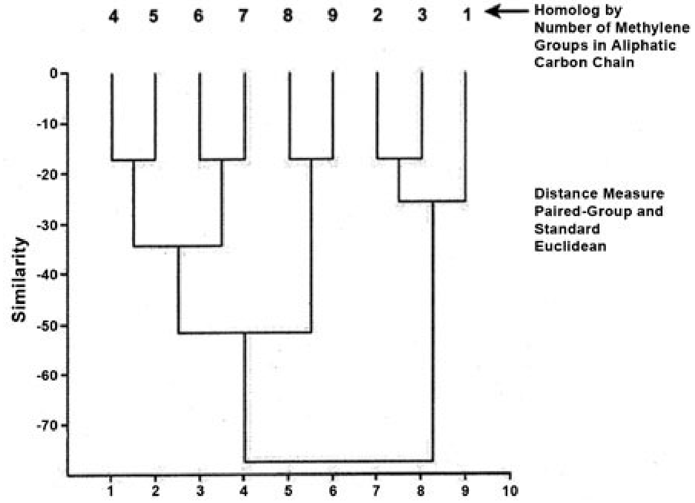

Figure 3. Cluster analysis (paired group, Euclidean distance) using properties in Table 1 indicate homologs having 4 and 5 methylene groups to be most similar. Similar pairs are 6 and 7, 8 and 9, with 2 similar to 3.

hierarchical classification of homologs based on properties of Table 1. The horizontal dendrogram shows classification into clusters which implies similarity among the subjects. Homologs 1, 2, and 3 fall within supercluster at node A, and are further classified with homologs 2 and 3 associated. Homologs 4 to 9 fall within supercluster at node B, which is divided further to supercluster at node C having homologs 6 to 9. Homologs 4 and 5 are most associated, followed by designated pairs 6 and 7, pairs 8 and 9. Again, homologs having the highest molecular weights (4, 5, 6, 7, 8, and 9) are distinguished from those of small molecular weights (1, 2, and 3). Subdivisions within superclusters associated the homologs into pairs with a series member of closest similarity. These types of analysis will allow determination of clinical drugs of most similar (or most dissimilar) activity and properties.

Cluster analysis is also referred to as segmentation analysis or taxonomy analysis.[27] The goal is to identify sets of groups (clusters), which both minimizes within-group variation and maximizes between-group variation. Hence subjects placed in the identical cluster have highest similarity. Figure 3 presents results of cluster analysis of homolog properties in Table 1 as a vertical dendrogram. Standard Euclidean (the shortest distance measured between 2 points) and paired-group distance (clusters are joined based on the average distance between members in 2 clusters) parameters are used. Homologs are identified by the number of methylene groups in aliphatic chain. Homologs are paired with another member of the series with the exception of homolog 1, which is seen as distinct but falling within a supercluster with homologs 2 and 3 (see Figure 3). Pairing occurred as follows: 4 and 5, 6 and 7, 8 and 9, 2 and 3. The 2 members of each pair are seen as more similar than for other pair clusters. This analysis will assist clinicians to identify drugs having greatest similarity or dissimilarity.

MDS, considered as an alternative to factor analysis, will detect underlying dimensions so that observed similarities (or dissimilarities) among subjects can be visualized (ie. As a 2-way plot) in a manner that preserves these distances.[28] Subjects that are most similar will be organized in a 2-way plot, so that they are in closest proximity. Figure 4 presents the results of nonmetric MDS of properties in Table 1. There are 2 forms of MDS, metric and nonmetric, in which nonmetric MDS requires the data to be in ranks only and has fewer restrictions than metric MDS. Results in Figure 4 clearly show homolog 1 (by number of methylene groups, see far right-hand side) to be distinct from the remaining members of the series. Homologs 2 and 3 are in closest proximity (see inset oval) and are considered by this algorithm to be most similar. Homologs 4, 5, and 6 are grouped and most similar (see inset rectangle), with homologs 7, 8, and 9 grouped. Again there is a distinction made among the highest molecular weight members in the series from the lower molecular weight homologs. Furthermore, this algorithm can distinguish as least 4 groupings (homolog 1 by itself) within the series.

Additional pharmacological properties are presented in Table 3, inclusive of drug likeness determined by 2 different methods. Identified by number of (-CH_2-) groups in the aliphatic chain, homologs 1 to 8 show zero violations of the Rule of 5, which indicates that these homologs will have favorable bioavailability.[29] The Rule of 5 states that a drug may show poor permeation when (1) formula weight is > 500; (2) Log P > 5; and (3) there are >10 H-bond

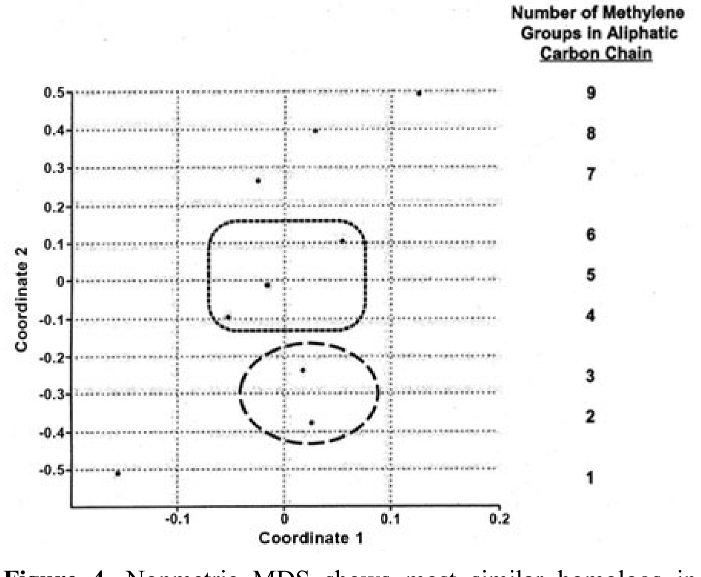

Figure 4. Nonmetric MDS shows most similar homologs in closest proximity. Using properties of Table 1: homologs 2 and 3 most similar (circle); homologs 6, 5, and 4 most similar (rectangle); with 9, 8, and 7 closest together.

Table 3. Molecular Properties of Homologous Series*

Number of Methylene Groups in Aliphatic Chain	Violations of Rule of 5	Solubility mg/L	Kp cm/h	Number of Oxygen and Nitrogen Molecules	Drug Likeness	
					Actelion	MolSoft
1	0	1659	0.00113	2	4.16	0.39
2	0	522.8	0.00207	2	4.17	0.84
3	0	164.5	0.00380	2	4.19	-0.58
4	0	51.63	0.00697	2	2.97	-0.80
5	0	16.18	0.0128	2	1.09	-0.80
6	0	5.064	0.0234	2	-1.24	-0.80
7	0	1.582	0.0429	2	-1.24	-0.80
8	0	0.4938	0.0786	2	-1.24	-0.80
9	1	0.1539	0.144	2	-1.24	-0.80

*Kp indicates dermal permeability coefficient Violations of Rule of 5 and number of oxygen and nitrogen molecules determined by Molinspiration cheminformatics. Solubility and dermal permeability coefficient Kp are determined by EpiSuite.

acceptors. Homolog shows a miLog P value of 5.404 and therefore may be sequestered in fatty tissue. Zero violations of Rule of 5 is a significant beneficial parameter and supports the contention of having clinical efficacy.[29] The number of methylene groups in aliphatic chain (see Figure 1) correlates highly (r = 0.8510) with numerical values of dermal permeability coefficient (Kp). Correlation of Kp to Rule of 5 is high (r = 0.8503). Solubility is inversely correlated to number of methylene groups (r = −0.7139) as is drug likeness determined by Actelion (r = −0.9345) and MolSoft (r = −0.7395). Aqueous solubility decreases as the number of methylene groups (and consequently Log P) increases; however, greater lipophilicity enhances dermal penetration, and the values of Kp increase and methylene groups increase. Drug likeness, defined by method of Actelion, is acquired if calculated values fall in the range of approximately −12 to 8. Consequently all homologs are identified as having actual drug likeness. Similarly, MolSoft defines drug likeness as a range of −2.00 to 2.3, in which all homologs are readily inclusive. Therefore, by 2 separate methods, the homologs of this series are readily identified as having good drug likeness.

Correspondence analysis is a method of factoring categorical variables and displaying them so that associations can be studied in 2 or more dimensions (ie, A 2-way plot).[30] Routine correspondence analysis can suffer from 2 problems: (1) arch effect owing to unimodal distribution; and (2) compression of data at initial and terminal ends. Detrending removes the arch effect and compression of data. Detrended correspondence analysis is performed on properties presented in Table 1, with homologs identified by the number of methylene groups in the aliphatic chain (see right-hand side Figure 5). A very mild curvature is observed with plotted subjects that are in sequential order 1 to 9 moving toward the origin. Such association indicates significant linear relationships within the data matrix (molecular properties of Table 1). Linear associations of properties listed in Table 1 are readily seen (indicated by Pearson correlation coefficient) and therefore correlate well with results of detrended correspondence analysis.

Multiple regression analysis can determine that a set of independent variables explains a significant proportion of variance (shown by R^2) in a dependent variable. The equation takes the following form:

$$y = b_1x_1 + b_2x_2 + \dots\ b_nx_n + c, \qquad (2)$$

where b values are regression coefficients, and c is the constant where the regression line intercepts the y-axis. Multiple regression analysis using properties of Table 1 showed several equations relating formula weight (FW) to descriptors molar volume (MV), molar refractivity (MR),

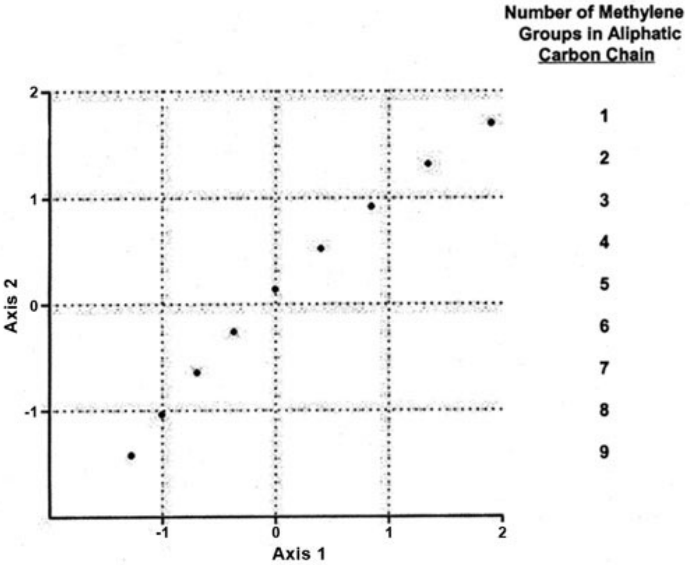

Figure 5. Detrended correspondence analysis removes the arch effect observed in routine correspondence analysis. Using properties in Table 1, homologs are ordered from 1 to 9, suggesting linear-type association in important properties.

parachor (PARACHOR), index of refraction (IndofRef), and number of rotatable bonds (nRotBonds). This equation becomes

$$FW = 157.046 - 0.0186\ (MV) + 0.0273\ (MR) + 0.0027\ (PARACHOR) \\ -0.7273\ (IndofRef) + 14.0982 \qquad (3)$$

Another form can be written using the number of rotatable bonds descriptor:

$$FW = 157.05 - 0.01864\ (MV) + 0.02727\ (MR) + 0.002727\ (PARACHOR) \\ -0.7273\ (IndofRef) + 14.098\ (nRotBonds) \qquad (4)$$

A reduced form, which uses molar volume, molar refractivity, and parachor appears as follows:

$$FW = 72.577 + 0.8613\ (MV) + 2.760\ (MR) + 1.388\ (PARACHOR) \qquad (5)$$

These regression equations may predict similar structures when property values are inserted as an independent or dependent variable.

Studies using multivariate methods have successfully elucidated properties for improvement of various drug delivery and dosage forms.[31-33]

The aliphatic chain imparts useful effects on molecular properties that have been revealed here through application of pattern recognition analysis. Enhanced blood-brain barrier penetration is shown through PSA, Log BB, and nonionization at pH 7.4 in the blood stream. Positioning of the N-mustard substituents achieved by varying the length of the aliphatic chain beneficially affects the properties of these homologs, a common goal of designing identical twin drugs.[34] Another benefit derived is the delivery of multiple active medicinal sites per drug molecule,[34] a result achieved with this homologous series as each member carries 2 bifunctional N-mustard groups for a total of 4 sites for alkylation (see Figure 1). Previous work has proven the efficacy of this approach for delivery of antibiotics,[35,36] Nonsteroidal antiinflammatory drugs (NSAIDs), diuretics, β-blockers, enzyme inhibitors, opiates,[34] and anticancers.[19]

The biopharmaceutical classification system (BCS) was developed primarily as a scientific framework to classify drug substances based on their aqueous solubility and intestinal permeability.[37,38] The BCS enables regulatory bodies to simplify and improve the drug approval process.[38] According to the BCS the following criteria are used: Class 1, high solubility-high permeability; Class 2, low solubility-high permeability; Class 3, high solubility-low permeability; and Class 4, low solubility-low permeability. Using these criteria and properties of the homologs including Log P, Kp, and aqueous solubility, it was determined that all 9 homologs fall into Class 2 (low solubility-high permeability). This result suggests that these homologs are suitable for application with the following drug delivery systems: micronization, lyophilization, addition of surfactants, emulsions, and use of complexing agents.

CONCLUSIONS

Properties of formula weight, molar volume, molar refractivity, number of rotatable bonds, and parachor are highly correlated and directly proportional to the number of methylene groups within the aliphatic carbon chain of the homologs, while the PSA remains constant at 6.476 A^2. This value of PSA indicates that all homologs will readily penetrate the blood-brain barrier and have greater than 95% absorption from the intestinal tract, while calculated values of partition coefficient Log P indicate homologs 1 and 2 (by number of methylene groups in aliphatic chain) will readily penetrate the central nervous system (Log P values indicate homologs 1, 2, and 3 are suitable for oral administration; homologs 4 and 5 are suitable for transdermal administration; and homolog 9 is suitable for sublingual administration). SOTA analysis, K-means cluster analysis, and discriminant analysis of molecular properties in Table 1 clearly distinguished the high molecular weight homologs from the low molecular weight homologs, while cluster analysis (paired-group distance) and hierarchical classification distinguished homologs by molecular weight, identified detailed intraseries similarities, and grouped members into pairs. Nonmetric MDS was able to associate homolog members of high similarity into the following groups: (A) homolog 1; (B) homologs 2 and 3; (C) homologs 4, 5, and 6: (D) homologs 7, 8, and 9.

ACKNOWLEDGMENTS

This work was funded by the College of Arts and Sciences, Chemistry Department, University of Nebraska, Omaha, NE 68182.

REFERENCES

1. Gareth T. *Medicinal Chemistry.* New York, NY: John Wiley and Sons; 2000.

2. Silverman R. *The Organic Chemistry of Drug Design and Drug Action*. San Diego, CA: Academic Press; 1992.

3. Pratt W, Ruddon R, Ensminger W, Maybaum J. *The Anticancer Drugs.* New York, NY: Oxford University Press; 1994.

4. Gringauz A. *Introduction to Medicinal Chemistry.* New York, NY: Wiley-VCH; 1997.

5. King F. *Medicinal Chemistry Principles and Practice.* Cambridge, UK: Royal Society of Chemistry; 2001.

6. Pain A, Samanta S, Dutta S, et al. Synthesis and evaluation of substituted naphthalimide nitrogen mustards as rationally designed anticancer compounds. *Acta Pol Pharm.* 2003;60:285–291.

7. Pain A, Samanta S, Dutta S, et al. Evaluation of napromustine, a nitrogen mustard derivative of naphthalimide, as a rationally

designed mixed-function anticancer agent. *Exp Oncol.* 2002;24: 173–179.

8. Faissat L, Martin K, Chavis C, Montero J, Lucas M. New nitrogen mustards structurally related to (L)-Carnitine. *Bioorg Med Chem.* 2003;11:325–334.

9. Fousteris MA, Koutsourea AI, Arsenou ES, Papageorgiou A, Mourelato D, Nikolaropoulos SS. Antileukemic and cytogenetic effects of modified and non-modified esteric steroidal derivatives of 4-methyl-3-bis(2-chloroethyl)amino benzoic acid (4-Me-CABA). *Anticancer Res.* 2002;22:2293–2299.

10. Papageorgiou A, Nikolaropoulos S, Arsenou E, et al. Enhanced cytogenetic and antineoplastic effects by the combined action of 2 esteric steroidal derivatives of nitrogen mustards. *Chemotherapy.* 1999;45:61–67.

11. Baraldi P, Romagnoli R, Giovanna P, Nunez M, Bingham J, Hartley J. Benzoyland cinnamoyl nitrogen mustard derivatives of benzoheterocyclic analogues of thetallimustine: synthesis and antitumor activity. *Bioorg Med Chem.* 2002;10:1611–1618.

12. Chen Z, Wan W, Xu D. Studies on the synthesis of tetrapyrrole nitrogen mustards and their directed dual anti-tumor activities. *J Chin Pharm Sci.* 1998;7:230–231.

13. O'Conner C, Denny W, Fan J, Gravatt G, Grigor B, McLennan D. Hydrolysis and alkylating reactivity of aromatic nitrogen mustards. *J Chem Soc.* 1991;12:1933–1939.

14. Palmer B, Wilson W, Cliffe S, Denny W. Hypoxia-selective antitumor agents. 5. Synthesis of water-soluble nitroaniline mustards with selective cytotoxicity for hypoxic mammalian cells. *J Med Chem.* 1992;35:3214–3222.

15. Catsoulacos P, Camoutsis C, Pelecanou M. Antileukemic activity of homo-azasteroidal esters of the isomers of N,N-bis(2-chloroethyl) aminocinnamic acid. *Eur J Med Chem.* 1991;26:659–661.

16. Hartley J, Preti C, Wyatt M, Lee M. Design, synthesis and biological evaluation of benzoic acid mustard derivatives of imidazole-containing and C-terminal carboxamide analogs of distamycin. *Drug Des Discov.* 1995;12:323–335.

17. Brooks N, McHugh P, Lee M, Hartley J. The role of base excision repair in the repair of DNA adducts formed by a series of nitrogen mustard-containing analogues of distamycin of increasing binding site size. *Anticancer Drug Des.* 1999;14:11–18.

18. Kovalenko S. Effect of basicity and details of chemical structure on the mutagenic activity of nitrogen mustards. *Genetika.* 1972; 8:100–103.

19. Bartzatt R, Donigan L. Two identical twin nitrogen mustard agents that express rapid alkylation activity at physiological pH 7.4 and 37 °C. *Lett Drug Des Discov.* 2004;1:78–83.

20. Herrero J, Valencia A, Dopazo J. A hierarchical unsupervised growing neural network for clustering gene expression patterns. *Bioinformatics.* 2001;17:126–136.

21. Palm K, Stenberg P, Luthman K, Artursson P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm Res.* 1997;14:568–571.

22. Ertl P, Bernhard R, Selzer P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J Med Chem.* 2000;43: 3714–3717.

23. Clark D. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 2. Prediction of blood-brain barrier penetration. *J Pharm Sci.* 1999;88:815–821.

24. Hansch C, Leo A, Hockman D. *Exploring QSAR: Hydrophobic, Electronic, and Steric Constants. ACS Professional Reference Book.* Washington, DC: The American Chemical Society; 1995.

25. Johnson R, Wichern D. *Applied Multivariate Statistical Analysis*. Englewood Cliffs, NJ: Prentice Hall Inc; 1992.

26. Dundteman G. *Introduction to Multivariate Analysis*. Beverly Hills, CA: Sage Publication; 1994.

27. Anderberg M. *Cluster Analysis for Applications*. San Diego, CA: Academic Press; 1973.

28. Schiffman S, Reynolds M, Young F. *Introduction to Multidimensional Scaling: Theory, Methods, and Applications*. New York, NY: Academic Press; 1981.

29. Lipinski A, Lombardo F, Dominy B, Feeney P. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 1997;23:3–25.

30. Greenacre MJ. *Correspondence Analysis in Practice*. London, UK: Academic Press; 1993.

31. Hardy I, Cook W. Predictive and correlative techniques for the design, optimization and manufacture of solid dosage forms. *J Pharm Pharmacol.* 2002;55:3–18.

32. Mareno M, Magalhaes N, Cavalcanti C, Alves A. Hierarchical cluster analysis applied to drug design. *Quim Nova.* 1996; 19:594–599.

33. Fossler M, Chang K, Young D. The use of cluster analysis in pharmacokinetics. *Acta Pharm Jugosl.* 1990;40:225–236.

34. Wermuth CG. *The Practice of Medicinal Chemistry.* San Diego, CA: Academic Press; 1996.

35. Aboul-Fadl T, Mahfouz N. Metronidazole twin ester prodrugs. 2. Non identical twin esters of metronidazole and some antiprotozoal halogenated hydroxy-quinoline derivatives. *Sci Pharm.* 1998;66:309–324.

36. Mahfouz N, Aboul-Fadl T, Diab A. Metronidazole twin ester prodrugs: synthesis, physiochemical properties, hydrolysis kinetics and antigiardial activity. *Eur J Med Chem.* 1998;33:675–683.

37. Rinaki E, Valsami G, Macheras P. Quantitative biopharmaceutics classification system: the central role of dose/solubility ratio. *Pharm Res.* 2003;20:1917–1925.

38. Lobenberg R, Amidon G. Modern bioavailability, bioequivalence and biopharmaceutics classification system: new scientific approaches to international regulatory standards. *Eur J Pharm Biopharm.* 2000; 50:3–12.