2003

Utilizing Structure Property Correlations to Predict and Analyze Two Derivatives of an Ampicillin Homologous Series

C. Malesa
University of Nebraska at Omaha

J. Stoddard
University of Nebraska at Omaha

Ronald Bartzatt
University of Nebraska at Omaha, rbartzatt@unomaha.edu

Journal of Undergraduate Chemistry Research, **2003**, 3, 127

UTILIZING STRUCTURE PROPERTY CORRELATIONS TO PREDICT AND ANALYZE TWO DERIVATIVES OF AN AMPICILLIN HOMOLOGOUS SERIES

C. Malesa*, J. Stoddard*; R. Bartzatt†

***University of Nebraska Biology Department 6001 Dodge Street Omaha, Nebraska 68182 USA**

†University of Nebraska, Durham Science Center, Chemistry Department, Medicinal Chemistry Laboratory, 6001 Dodge Street, Omaha, NE, 68182

Abstract

Structure Property Correlation methods such as regression analysis and pattern recognition are applied to predict the molecular properties of two members of an ampicillin homologous series. The pentyl and hexyl esters of ampicillin are also evaluated for their similarity to other penicillins by use of multiple regression, contingency tables, cluster analysis, correspondence analysis, self organizing tree algorithms, factor analysis, principal component analysis, box plots, and other graphing methods. Other members of the homologous series include methyl, ethyl, propyl, and butyl esters of ampicillin which have been previously synthesized and tested in tissue culture against *Escherichia coli*. All of the tested esters of ampicillin significantly inhibited penicillin susceptible and ampicillin resistant bacteria, as well as streptomycin resistant bacteria. Drug homologous series has been observed with other antibiotics and medicinal compounds. Homologous series activity is a trend observed for ampicillin by graphing the reciprocal equi-effective concentrations versus the number of carbons comprising the ester group. Regression analysis and contingency tables predict the molecular properties of the hexyl and pentyl esters, while cluster analysis, factor analysis, correspondence analysis, principal component analysis, and tree algorithms correlate them to the parent ampicillin and other members of the β-lactam class of antibiotics. This work demonstrates the effectiveness of applying numerical analysis methods to drug design and development.

Keywords: ampicillin, multiple regression, cluster analysis, pattern recognition.

Introduction

Ampicillin is a broad spectrum antibiotic utilized clinically against gram-negative bacteria (1). It is a semi-synthetic drug which interferes with the formation of the bacterial cell wall (1,2). The β-lactam ring characteristic of penicillin class antibiotics is mistaken by transpeptidase for the terminal end -D-alanine-D-alanine sequence which this enzyme bonds to either a pentaglycine group or other peptides to seal the bacterial cell wall (2). Ampicillin has a carboxylic acid group as part of its structure which serves as an effective site for prodrug and derivative modifications. Previous studies have shown that the ester derivatives (ie. methyl, ethyl, propyl, and butyl) have significant antibacterial action targeting penicillin susceptible and resistant *Escherichia coli*, and streptomycin resistant bacteria (3,4,5). A homologous series are drugs which differ from each other by only a $-CH_2$ group (6). It has been shown previously that homologous series of antibiotics have increased pharmacological activity as the number of carbons in the ester group increases from 1 to 7 carbons (7). The carbon lengthening must be aliphatic and not have any branches (7). The usual shape of an activity versus carbon length plot has an initial ascending branch (carbon 1 to 6 or 7), a single maxima (at the 6 or 7 carbon), and a descending branch (more than 7 carbons) (6,7).

The general purpose of multiple regression is to analyze the relationship between several independent or predictor variables and dependent or criterion variables. The term cluster analysis encompasses a number of different classification algorithms which can be used to develop taxonomies (8,9). Correspondence analysis is a descriptive/ exploratory method designed to analyze simple two-way and multi-way tables containing some measure of correspondence between the rows and columns, giving results similar to factor analysis (10,11,12). Principal components analysis is a linear dimensionality reduction technique and projects the data into a lower-dimensionality space (ie. fewer variables) (13,14). Box plots show ranges or distribution characteristics of values of a selected variable (or variables) and central tendency (ie. Median or mean) with range or variation statistics

(15,16). Any outlier data points can also be visualized by box plots. Contingency tables represent two sets of variables and can predict numerical values similarly to multiple regression (12,14).

These numerical techniques are combined to predict the molecular properties of the hexyl and pentyl ester derivatives of ampicillin. Then methods such as factor analysis (FA), principal component analysis (PCA), cluster analysis (CA), correspondence analysis (CA), and self organizing tree algorithm (SOTA) (17) will be utilized to confirm and verify the similarity of the pentyl and hexyl ampicillin esters to other members of the penicillin class of antibiotics. This work demonstrates the following: 1) Adherence of the methyl, ethyl, propyl, butyl, pentyl, and hexyl esters of ampicillin as members of a homologous series; 2) The ability of multiple regression methods to predict the molecular properties of the pentyl and hexyl ampicillin esters; 3) Show the similarities of the pentyl and hexyl esters to other clinically utilized penicillin drugs; and 4) Demonstrate the efficacy of pattern recognition methods for drug design.

Experimental

Software and Algorithms

An Intellistation E Pro IBM with Windows Professional was utilized for statistical analysis and plotting. Quattro Pro 10 was utilized for graphing. Self organizing tree algorithm and cluster analysis accomplished by the GEPAS, SOM PAK, Self Organizing Map Version 3.1 (Helsinki University of Technology Laboratory of Computer and Information Science Rakentajanaukio 2 C, SF-02150 Espoo FINLAND)(17). Multiple regression accomplished by Quattro Pro 10 and Orlando Mansur Numerical Mathematics (PO Box 16166, Jersey City, NJ 07306, USA). Cluster analysis, principal component analysis, and correspondence analysis accomplished by AMADA v.2.0.3 (copyright X. Xia and Z. Xie (18)) and KyPlot v.2.0 beta 15 (copyright Koichi Yoshioka). Contingency tables, multiple regression, and factor analysis accomplished by KyPlot. Molecular properties of antibiotics were calculated by Chemsilico LLC, Tewksbury, MA 01876) and ChemSketch (ACD/ ChemSketch, 90 Adelaide Street West, Toronto, Ontario, M5H 3V9 Canada). Molecular modeling was accomplished by ChemSketch. Polar surface area and formula weight (based on structure) were calculated by Molinspiration java (Molinspiration, Liscie udolie 2, SK-841 04 Bratislava, Slovak Republic).

Results and Discussion

Ampicillin is a broad-spectrum antibiotic and one of the most widely used drugs. The rise of ampicillin resistant bacteria is of great concern and development of new derivatives of the β-lactam class of antibiotics is a continuing process (2). The molecular structures of the ampicillin parent compound and its esters are shown in [Figure 1.]. Note that the addition of methylene groups (-CH_2-) form an aliphatic chain with no branching. Ampicillin has a carboxyl group which may ionize in alkaline conditions such as the blood at pH 7.4 producing a species with good aqueous solubility but poor lipid solubility (Log P = -0.17). Ionization of ampicillin will increase the rate of elimination due to its greater water solubility, which decreases the length of time of medicinal activity in the blood stream (2,3). The Log P of the ester derivatives of ampicillin are methyl (0.37), ethyl (0.95), propyl (1.57), and butyl (2.16). The greater numerical positive values indicate drugs of greater lipid solubility (ie. solubility in cell membranes) with the benefit of longer survival in vivo and therefore greater medicinal benefit. These pharmacological traits are an example of the benefit of novel drug design strategies.

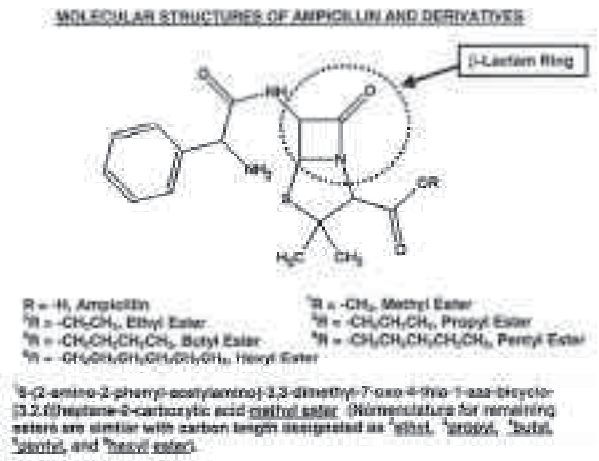

Figure 1: Molecular structures of the parent compound ampicillin and the ester derivatives.

Important molecular properties formula weight (FW), molar volume (MV), molar refractivity (MR), parachor (PARA), polar surface area (PSA) were calculated via ChemSketch for all four esters of ampicillin and other β-lactam antibiotics (carbenicillin, penicillin K, penicillin X, methicillin, penicillin G, penicillin F, penicillin dihydro F). These values constitute properties or descriptors of these drugs which are presented in Table 1 and will comprise the training set of data for multiple regression prediction and pattern recognition by the methods described in Experimental. The properties themselves describe the steric and van der Waals characteristics of the antibiotic.

Presented in [Figure 2.] are plots of the molecular property values for antibiotics in Table 1. Graphing is accomplished by Quattro Pro 10, with equations of the lines formed and Pearson correlation coefficients shown in the figure (x-axis as formula weight and the independent variable). Utilizing the multiple regression equations the values of molecular properties for the pentyl and hexyl esters of ampicillin can be calculated and are shown as follows with values estimated by ChemSketch (see Experimental) for comparison (multiple regression/ ChemSketch): **PARACHOR**=(pentyl-862.6/906.4), (hexyl-891.5/946.4); **MOLARVOLUME** = (pentyl-303.6/328.8), (hexyl-313.4/345); **MOLAR REFRACTIVITY**=(pentyl-109/113.3), (hexyl-112.9/ 117.9); **POLAR SURFACE AREA** = (pentyl-113.9/ 101.74), (hexyl-117.3/101.74). Differences in the values range from 7.6% in molar volume to about 4% for molar refractivity. These multiple regression results stem from a single independent variable (formula weight) and four dependent variables PARA, MV, MR, and PSA.

Another approach utilizing multiple regression is applying several independent variables (PARA, MV, MR, PSA) and fewer dependent variables (formula weight) to reduce relative error. Initially it requires the estimation of molecular properties by software which can evaluate the numerical values based upon a submitted molecular structure, which is accomplished here by ChemSketch. Estimation of the molecular properties come from summations of contributions by structure fragments. A linear regression method by Mansur (see Experimental) produces an equation that incorporates PARA, MV, MR, and PSA as X values and generates the

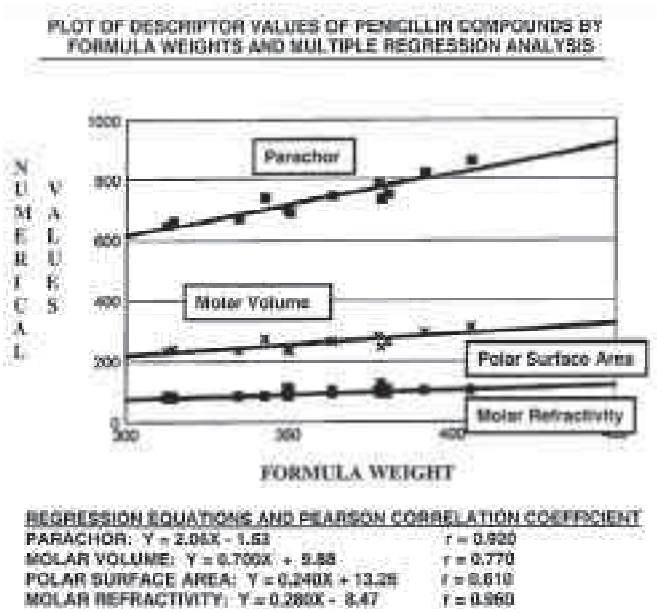

Figure 2: Linear results obtained after multiple regression of molecular properties for all β-lactam antibiotics (inclusive of the pentyl and hexyl esters of ampicillin).

dependent Y value as formula weight. The equation appears as follows:
Y = 18.88 - (1.5512)(MV) - (0.5956)(MR) + (1.088)(PARA) - (0.03589)(PSA). Pearson correlation coefficient is 0.9909 with formula weights for the pentyl ester at 424.2 and the hexyl ester at 439.8 (actual values are 419.54 and 433.57, respectively). Relative error in FW values are only 1.3% and 1.4%, respectively. It is easily seen that prediction of a single dependent property is enhanced by the use of several independent properties. This is a useful concept when predicting the properties of potential but not yet synthesized drugs.

It has been shown in previous studies that the esters of ampicillin inhibit bacteria. The medicinal activity of these esters were calculated and compared to the number of carbons comprising the aliphatic branch of the ester (ie. -O-R), see [Figure 3.]. The equation for the ascending branch is shown in inset as y = 0.447Lnx + 0.6393. The medicinal activity is expressed as Equi-effective Concentration (EEC), where EEC is defined as concentration of derivative drug at 50% activity divided by the concentration of parent drug at 50% activity, or EEC = IC_{50} test/ IC_{50} parent. Values of IC_{50} were determined in previous work (3,4,5) and are as follows: Methyl ester= 1.75 mg/mL; Ethyl ester= 0.877 mg/mL; Propyl ester= 1.0 mg/mL; Butyl ester= 0.789 mg/ mL. The data comprises an ascending trend

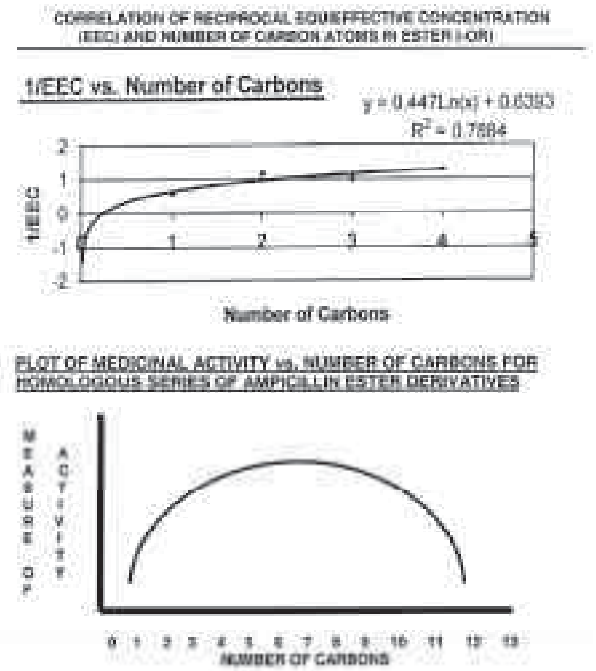

Figure 3: Plot of l/EEC versus number of carbons in ester aliphatic chain (-O-R) shows adherence to rule of homologous series. The general trend of medicinal activity for a homologous series is presented in the lower half.

following the configuration of the expected activity of a homologous series shown in [Figure 3.] lower half. Note the ascending branch (1 to 5 carbons), a maxima (6 to 7 carbons), and descending branch (8 to 12 carbons). This shows clearly that the rule of homologous series is obeyed and it would follow that the next higher members of the series, the pentyl and hexyl esters, will be even more effective in medicinal activity. The clinical potential of the higher series members, pentyl and hexyl esters, would be significant and beneficial.

Graphical methods can provide a variety of configurations to represent the data for visual inspection [Figure 4.] shows two such representations as a means plot and box plot. ties of all the penicillin drugs presented in Table 1 with the addition of the pentyl and hexyl ampicillin esters are graphed in [Figure 4.]. Note that there is tight clustering of the data points around the mean of their respective property (FW, MR, MV, PARA, and PSA). The inset arrow designates a data point of greater distance from the mean of Parachor values but is not an outlier, as determined by administering the Grubbs Test for detecting outlying observations (19). The boxplot representation also indicates this data point (see inset arrow in boxplot). This visualization supports the contention that the pentyl and hexyl esters are associated with and similar to other penicillin drugs of Table 1.

Additional numerical methods exist which can determine the similarity of the pentyl and hexyl esters to actual clinical penicillin drugs. This will confirm the results of the predictive multiple regression methods applied previously. These methods can be considered pattern recognition methods and include contingency tables, cluster analysis, SOTA analysis, principal component analysis, correspondence analysis, and factor

TABLE 1. DESCRIPTORS OF PENICILLIN ANTIBIOTICS

	FORMULA WEIGHT	MOLAR REFRACTIVITY	MOLAR VOLUME	PARACHOR	POLAR SURFACE AREA
CARBENICILLIN	**378.4**	**92.6**	**246.1**	**735.2**	**124.**
PENICILLIN K	**342.46**	**89.41**	**273.3**	**741.6**	**86.7**
PENICILLIN X	**350.39**	**87.87**	**232.**	**690.2**	**106.9**
METHICILLIN	**380.42**	**94.44**	**262.5**	**752.1**	**105.2**
PENICILLIN G	**334.39**	**86.34**	**235.1**	**675.**	**86.7**
PENICILLIN F	**312.39**	**80.06**	**234.2**	**649.**	**86.7**
PENICILLIN DIHYDRO F	**314.4**	**80.15**	**240.8**	**661.4**	**86.7**
AMPICILLIN	**349.41**	**89.94**	**239.3**	**702.7**	**112.7**
>Methyl Eater	**363.43**	**94.78**	**264.2**	**746.1**	**101.74**
>Ethyl Ester	**377.47**	**99.41**	**280.4**	**786.1**	**101.74**
>Propyl **Ester**	**391.49**	**104.04**	**296.5**	**826.2**	**101.74**
>Butyl Ester	**405.51**	**108.67**	**312.7**	**866.3**	**101.7**

Units for molar volume, molar refractivity, parachor cm^3
Units for polar surface area = $angstroms^2$

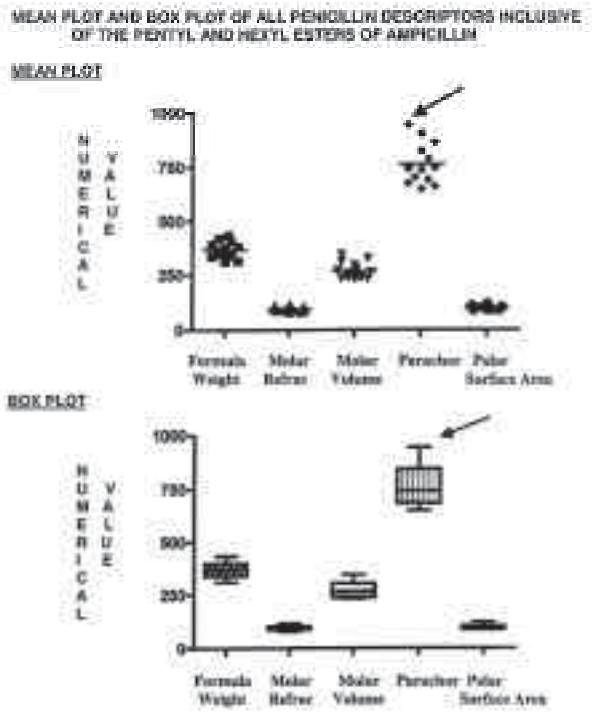

Figure 4: A means plot and box plot show clearly the close proximity of numerical values of molecular properties for FW, MR, MV, PARA, and PSA.

analysis. These methods applied to drug design and development will be discussed now.

Contingency tables are for cross-tabulated or cross sectional data and measure the degree of association of the rows with columns. Table 2 shows the results of such a correlation and prediction via KyPlot (see Experimental), which utilizes integer values for the molecular properties. Columns and rows are clearly associated by numerical values of predicted Properties. This result supports the multiple regression prediction and estimation values calculated by ChemSketch. This also supports the determination of the pentyl and hexyl esters of ampicillin as viable members of the homologous series and are similar to other clinical penicillin compounds. The Pearson correlation coefficients for comparing the actual values to the contingency table predicted values for pentyl and hexyl esters are 0.999 and 0.999, respectively. Pearson correlation coefficients for the remaining penicillin drugs are > 0.960.

TABLE 2 CORRELATION AND PREDICTION BY CONTINGENCY TABLE [ACTUAL/PREDICTED]

	FORMULA WEIGHT	MOLAR REFRACTIVITY	MOLAR VOLUME	PARACHOR	POLAR SURFACE AREA
CARBENICILLIN	378	93	246	735	124
	362.70	94.27	267.09	752.67	99.27
PENICILLIN K	342	89	273	742	87
	362.81	91.70	259.81	732.13	96.56
PENICILLIN X	350	88	232	690	107
	337.62	87.75	248.62	700.61	92.41
METHICILLIN	380	94	263	752	105
	366.84	95.35	270.14	761.26	100.40
PENICILLIN G	334	86	235	675	87
	326.10	84.76	240.15	676.73	89.26
PENICILLIN F	312	80	234	649	87
	313.45	81.47	230.83	650.47	85.79
PENICILLIN DIHYDRO F	314	80	241	661	87
	318.28	82.72	234.39	660.49	87.11
AMPICILLIN	349	90	239	703	113
	343.83	89.36	253.20	713.51	94.11
> Methyl Ester	363	95	264	746	102
	361.32	93.91	266.08	749.80	98.89
> Ethyl Ester	377	99	280	786	102
	378.35	98.34	278.62	785.14	103.55
> Propyl Ester	391	104	297	826	102
	395.84	102.88	291.50	821.44	108.34
> Butyl Ester	405	109	313	866	102
	413.10	107.37	304-21	857.26	113.07
> Pentyl Ester	419	113	329	906	102
	430.13	111.79	316.75	892.60	117.72
> Hexyl Ester	434	118	345	946	102
	447.62	116.34	329.63	928.89	122.51

Cluster analysis (CA) is a classification method that is used to arrange a data set into clusters. The goal is to establish a set of clusters such that data within a cluster are more similar to each other than they are to data in another cluster. A cluster analysis of penicillin properties (inclusive of the pentyl and hexyl esters in order to determine similarity) is presented in (Figure 5.]. The approach utilized is single linkage (closest neighbor), hierarchical (creating groups through dividing), and divisive in hierarchy (all observations begin in one group and are divided by steps, see [Figure 5.]). Initially all the drugs are grouped, then two major nodes are formed of which A is subdivided into secondary nodes B and C which includes all ester derivatives (drugs 9 to 14), ampicillin, carbenicillin, penicillin X, and methicillin (ie. these are similar drugs). Tertiary node E contains penicillin X and ampicillin, with tertiary node D containing methicillin with all the ester derivatives. These results show that the pentyl and hexyl esters of ampicillin are associated with penicillin drugs already in clinical use. Ampicillin is clustered with penicillin X under tertiary node E and sub-grouped under secondary node B with penicillins 4, 9, 10, 11, 12, 13, and 14.

SOTA analysis is a combination of neural network analysis and cluster analysis (17). Neural networks are modeling methods that are nonlinear. Neural networks learn by example, gather representative data in which they invoke training algorithms to automatically learn the structure of the data. SOTA is a combination of neural network analysis and cluster analysis that produces results in the form of clusters. The molecular properties of all penicillin compounds (inclusive of the pentyl and hexyl ester derivatives that are being compared) was inserted into the SOTA algorithm (complete linkage and standard euclidean distances) and produced the following clusters, which again shows that an association exists among ester derivatives of ampicillin with clinically applied penicillins. CLUSTER 1: ampicillin, carbenicillin, penicillin K, penicillin X, methicillin, penicillin G, penicillin F, penicillin dihydro F, and methyl ester of ampicillin; CLUSTER 2: the ethyl, propyl, butyl, pentyl, and hexyl esters of ampicillin.

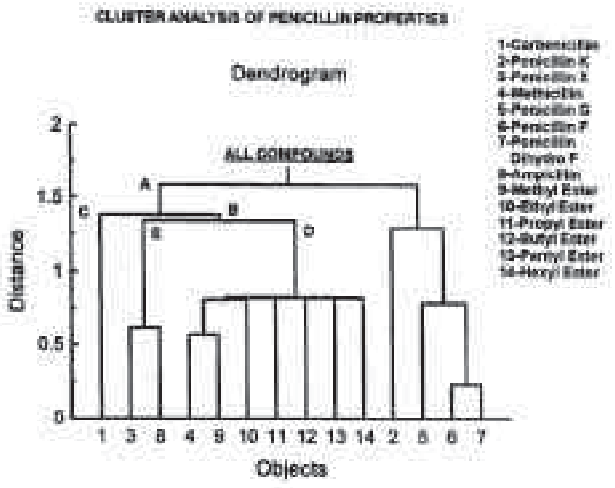

Figure 5: Dendrogram of cluster analysis for antibiotics 1 through 14 (see inset key to antibiotics identification).

The central concept of PCA is representation and summarization by combining two correlated variables into one factor. Therefore the number of variables are reduced and structure among the variables is ascertained. PCA can be made to produce a correlation coefficient matrix of the results of primary component mining. This was accomplished utilizing all penicillins of Table 1 with the pentyl and hexyl esters of ampicillin (utilizing ChemSketch values of properties), results shown in Table 3. High inter-column correlation was observed in general ($r > 0.870$, see Table 3), only polar surface area (PSA) showing low correlation with FW, MR, MV, and PARA. Remembering that pentyl and hexyl ester are included in the PCA correlation analysis, this shows strong inter-drug association and similarity. PCA is preferred as a method of data reduction while factor analysis is preferred to detect structure.

Correspondence analysis is similar to factor analysis. CA is a descriptive/exploratory method to analyze 2-way and multi-way tables showing association between rows and columns. Plots of the two dimensions of CA analysis of all penicillins (inclusive of pentyl and hexyl ester) is shown in [Figure 6.] as CA-1 versus CA-2. The proximities and absolute value of the distances of points relative to each other are important. Ampicillin esters are encased in the inserted rectangle and linear from left to right: methyl, ethyl, propyl, butyl, pentyl, and hexyl esters. Although they are up to 1.3 CA-1 units from the origin the other penicillin are well dispersed with even greater distances from the origin (penicillin K(inset circle) is > -1.5 CA-2 units away from origin). Carbenicillin is the greatest distance from the origin at more than -2 CA-1 units. Penicillin

G is > +1.7 CA-2 units above the origin (inset arrow). By visual inspection essentially all penicillin drugs, inclusive of esters of ampicillin, are sufficiently close in proximity to be considered similar.

Factor analysis is utilized to condense information into a smaller set of factors and analyze their interrelations for common underlying patterns. A plot of FA analysis will be analogous to correspondence analysis and are shown together in [Figure 6.]. Plotting the factors FA-1 and FA-2 in two dimensions shows the dispersion for each penicillin compound with proximities and distances again portraying associations. The inset rectangle encases ampicillin derivatives from left to right: methyl, ethyl, propyl, butyl, pentyl, and hexyl esters. Their overall distances from the origin are greater than generally observed with other drugs in this analysis (the furthest is the hexyl ester at +2 FA-1 units) however the methyl, ethyl, and propyl esters are within the distances calculated for penicillin K at > -1 FA-2 units (see inset arrow) and carbenicillin at > +2 FA-2 units (see inset circle). Visual inspection of these graphical results show clearly that the several esters of ampicillin can be considered associated with clinically applied penicillins.

TABLE 3 PRINCIPAL COMPONENT ANALYSIS BY KyPlot

Correlation Matrix of Penicillin Properties by Principal Component A

	FW	MR	MV	PARA	PSA
FW	1				
MR	0.975	1			
MV	0.885	0.952	1		
PARA	0.955	0.991	0.982	1	
PSA	0.482	0.319	0.071	0.245	1

HIGH CORRELATION

FORMULA WEIGHT AND MOLAR REFRACTIVITY
FORMULA WEIGHT AND PARACHOR
MOLAR VOLUME AND MOLAR REFRACTIVITY
MOLAR VOLUME AND FORMULA WEIGHT
PARACHOR AND MOLAR REFRACTIVITY
PARACHOR AND MOLAR VOLUME

LOW CORRELATION

POLAR SURFACE AREA AND FORMULA WEIGHT
POLAR SURFACE AREA AND MOLAR REFRACTIVITY
POLAR SURFACE AREA AND MOLAR VOLUME
POLAR SURFACE AREA AND PARACHOR

Conclusions

Similarities in molecular properties among the ester derivatives of ampicillin have been presented and utilized for comparison with clinical β-lactam antibiotics. Multiple regression methods showed by graphs and calculations that the higher members of the homologous series, pentyl and hexyl ampicillin ester, are linear related and highly correlated by properties. The properties of the pentyl and hexyl esters can be predicted by multiple regression and contingency table calculations. The methyl, ethyl, propyl, and butyl esters of ampicillin clearly form a homologous series when their medicinal activity is plotted as reciprocal equi-effective concentration (1/EEC) versus number of carbon atoms in the aliphatic non-branched ester group (-O-R). Mean plots and box plots show visually that property values of all penicillin drugs (inclusive of pentyl and hexyl esters) are closely grouped around their means, indicating association and similarity. Cluster analysis and

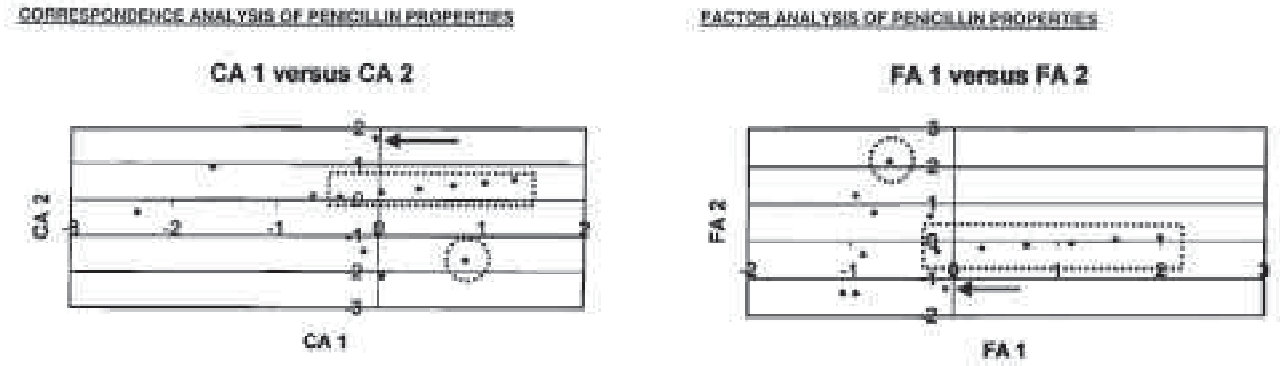

Figure 6: 2-D graphs depicting results of correspondence analysis analysis and factor analysis. For CA the ampicillin esters are encased in the inserted rectangle and linear from left to right: methyl, ethyl, propyl, butyl, pentyl, and hexyl esters. Arrows and circles other β-lactam antibiotics. For FA the ampicillin esters are within the inset rectangle from left to right: methyl, ethyl, propyl, butyl, pentyl, and hexyl esters.

SOTA analysis determined that the esters of ampicillin are associated with other β-lactams of the training matrix (see Table 1). Calculation of a correlation matrix by PCA showed clearly that properties of all the penicillins (inclusive of pentyl and hexyl esters) are highly correlated in formula weights, molar refractivity, molar volume, and parachor. CA and FA are similar methods which reduce raw data to fewer variables and looks for inter-relationships. Graphical representation of CA results showed clearly by inspection that molecular properties of all the penicillins are proximal and associated. FA results showed somewhat more dissimilarity among higher members of the homologous series of ampicillin with clinical penicillins. In summation, the results strongly support the contention that pentyl and hexyl esters of ampicillin are true members of a homologous series of ampicillin and will be effective and beneficial antibiotics.

Acknowledgments

This study was financially supported by the College of Arts and Sciences and the Chemistry Department of the University of Nebraska.

References

(1). B. Brock; M. Madigan. *Biology of Microorganisms*, Prentice Hall, Englewood Cliffs, New York, USA (1991) pp. 343-365.
(2). B. Davis, R. Dulbecco, H. Eisen; H. Ginsberg. *Microbiology*, J.B. Lippincott Co., New York, USA (1990) pp. 36-280.
(3). R. Bartzatt; C. Malesa. *Chemotherapy*, **2003**, 7(69), in press.
(4). R. Bartzatt; C. Malesa. *Biotechnol. App. Biochem.*, **2002**, 36, 89-93.
(5). R. Bartzatt, T. Benish, K. Koziol; J. Stoddard. *Physiological Chem. Phyl. & Med.*, **2000**, 32, 49-56.
(6). C.G. Wermuth. *The Practice of Medicinal Chemistry*, Academic Press, San Diego, USA (1996) pp. 182-192.
(7). R. Silverman. *The Organic Chemistry of Drug Design and Drug Action*, Academic Press, San Diego, USA (1992) pp. 15-21.
(8). J.H. Ward. *Journal of the American Statistical Association*, **1963**, 58, 236.
(9). J. Zupan. *Clustering of Large Data Sets*, Research Studies Press, New York, USA (1982).
(10). M.J. Greenacre. *Biometrica*, **1988**, 75, 457-467.
(11). H.H. Ku, R.N. Varner; S. Kulback. *Journal of the American Statistical Association*, **1971**, 66, 55-64.
(12). T.W. Anderson. *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, USA (1984).
(13). C.F. Gerald and P.O. Wheatley. *Applied Numerical Analysis*, Addison Wesley, Reading MA, USA (1989).
(14). T.W. Anderson. *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, USA (1958).
(15). W.S. Cleveland. *Graphs in Scientific Publications*, **1984**, 38, 270-280.
(16). W.S. Cleveland. *The Elements of Graphing Data*, Wadsworth, California, USA (1985).
(17). J. Herroero, A. Valencia; J. Dopazo *Bioinformatics*, **2001**, 17, 125-136.
(18). X. Xia; Z. Xie. *Bioinformatics*, **2001**, 17, 569-570.
(19). F. Grubbs. *Technometrics*, **1969**, 11(1), 1-21.