Equations A6 and A7 are rearranged to yield Eqs. A8 and A10 in the Laplace domain and Eqs. A9 and A11 in the time domain:

$$R(s) = Q_D(s)_i - (m_1 + m_2)s^{-1}R(s) - (m_1m_2)s^{-2}R(s) \quad \text{(Eq. A8)}$$

$$R(t) = Q_D(t)_i - (m_1 + m_2)\int R(t)\,dt - m_1m_2\int [\int R(t)\,dt]\,dt \quad \text{(Eq. A9)}$$

$$Q_B(s)_i = (A_1 + A_2)s^{-1}R(s) + \alpha s^{-2}R(s) \quad \text{(Eq. A10)}$$

$$Q_B(t)_i = (A_1 + A_2)\int R(t) + \alpha\int [\int R(t)\,dt]\,dt \quad \text{(Eq. A11)}$$

Equation A9 is solved on the analog computer to give $R(t)$; Eq. A11 is then solved to give $Q_B(t)$ using $R(t)$ and its integrals as shown by the circuit diagram in Fig. 9.

A similar approach can be applied to solve for input functions by rearranging Eqs. A6 and A7 to give $R(s)$ and the input function and solving these on the analog computer. Therefore, both convolution and deconvolution can be performed by this general approach which, when modified in this manner, can also be used to deconvolve $Q_D(t)_r$ against $Q_B(t)_r$ to obtain $G_{DB}(t)_r$.

REFERENCES

(1) S. A. Kaplan, in "Dissolution Technology," L. F. Leeson and J. T. Carstensen, Eds., APhA Academy of Pharmaceutical Sciences, Washington, D.C., 1974, pp. 163–187.

(2) V. F. Smolen, P. B. Kuehn, and E. J. Williams, *Drug Develop. Commun.*, **2**, 143 (1975).

(3) V. F. Smolen and W. A. Weigand, *J. Pharm. Sci.*, **65**, 1718 (1976).

(4) V. F. Smolen, *ibid.*, **60**, 878 (1971).

(5) V. F. Smolen, *Hosp. Pharm.*, **4**, 14 (1969).

(6) W. Brownell, S. Riegelman, and W. J. Mader, "Committee Report on Drug Dissolution Methodology," APhA Academy of Pharmaceutical Sciences, Washington, D.C., 1968.

(7) G. Levy and L. E. Hollister, *J. Pharm. Sci.*, **58**, 1368 (1969).

(8) G. Levy, *ibid.*, **52**, 1039 (1963).

(9) M. Gibaldi and H. Weintraub, *ibid.*, **58**, 1368 (1969).

(10) H. Weintraub and M. Gibaldi, *ibid.*, **59**, 1792 (1970).

(11) J. G. Wagner, P. G. Welling, P. L. Kwang, and I. E. Walker, *ibid.*, **60**, 666 (1971).

(12) W. A. Cressman, C. A. Janicki, P. C. Johnson, J. T. Doluisio, and G. A. Braun, *ibid.*, **58**, 1516 (1969).

(13) V. F. Smolen, *J. Pharmacokinet. Biopharm.*, **4**, 337 (1976).

(14) *Ibid.*, **4**, 355 (1976).

(15) D. S. Riggs, "Control Theory and Physiological Feedback Mechanisms," Williams & Wilkins, Baltimore, Md., 1970, pp. 23–26.

(16) *Ibid.*, pp. 91–97.

(17) E. R. Rodeman and J. T. P. Yao, "Structural Identification—Literature Review," NSF Technical Report Ce-STR-73-3, School of Civil Engineering, Purdue University, West Lafayette, Ind., 1973.

(18) C. D. McGillem and G. R. Cooper, "Continuous and Discrete Signal and System Analysis," preliminary ed., School of Electrical Engineering, Purdue University, West Lafayette, Ind., 1971, pp. 6–1 ff.

(19) R. D. Schoenwald, Ph.D. thesis, Purdue University, West Lafayette, Ind., 1971, p. 313.

(20) M. Pernarowski, in "Dissolution Technology," L. F. Leeson and J. T. Carstensen, Eds., APhA, Academy of Pharmaceutical Sciences, Washington, D.C., 1974, pp. 58–105.

(21) C. D. McGillem and G. R. Cooper, "Continuous and Discrete Signal and System Analysis," preliminary ed., School of Electrical Engineering, Purdue University, West Lafayette, Ind., 1971, pp. 3–43 ff.

(22) V. F. Smolen, E. J. Williams, and P. B. Kuehn, *Can. J. Pharm. Sci.*, **10**, 95 (1975).

(23) P. B. Kuehn, Ph.D. thesis, Purdue University, West Lafayette, Ind., 1974.

ACKNOWLEDGMENTS AND ADDRESSES

Received March 12, 1976, from the *Interdisciplinary Drug Engineering and Assessment Laboratory and the Department of Industrial and Physical Pharmacy, School of Pharmacy and Pharmacal Sciences, Purdue University, West Lafayette, IN 47907.*

Accepted for publication May 4, 1976.

[x] To whom inquiries should be directed.

Pattern Recognition II: Investigation of Structure–Activity Relationships

GOVIND K. MENON and ARTHUR CAMMARATA [x]

Abstract □ A simple form of pattern recognition is successfully used to classify a set of structurally diverse therapeutic agents. By using only organic structural information, the major pharmacological classes present were correctly identified and the pharmacologically unrelated compounds were separated out. One technique of factor analysis—principal component analysis—is shown to be readily adaptable in preprocessing the data. Simple graphical representation of the results enables their direct interpretation.

Keyphrases □ Structure–activity relationships—determined using pattern recognition methods of analysis, various drugs □ Pattern recognition—analysis methods used to determine therapeutic classes of various drugs □ Factor analysis techniques—used to determine therapeutic classes of various drugs

The application of pattern recognition methods in solving chemical problems has provided the incentive for considerable research within the past few years (1–6). Recently, these methods have been applied in the development of new biological agents due to their capacity to analyze rapidly large stores of accumulated information and to detect substances worthy of further investigation (7–10). In other words, the techniques can be directed toward establishing structural specificity in biological action, providing rationales in selecting substances for biological assay, and identifying pharmacophoric patterns of molecular substitution (1).

Several detailed accounts describe the various pattern recognition techniques (11–14), but a brief conceptual summary follows. Pattern recognition involves the detection and recognition of regularities or invariant properties among accumulated (often large) sets of measurements, the purpose being to provide a basis for new hypotheses. For example, consider an attempt to derive a

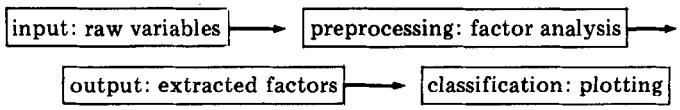
structure–activity relationship. If the relationship of structure to activity is to be developed from a basis involving a small number of structurally diverse compounds or from a large number of structurally similar compounds, then a visual examination is often sufficient. However, if a large number of structurally diverse compounds is involved, visual examination of the data to arrive at a structure–activity relationship is cumbersome if not impossible. Pattern recognition methods then can be employed to advantage.

A few reports in the recent literature describe the application of pattern recognition techniques in the investigation of structure–activity relationships (7–10). These have been concerned with two specific areas: classification of sedatives and tranquilizers (7, 9) and prediction of the activity of potential anticancer agents (8, 10). Although these investigations have been fairly successful, criticism (15, 16) has been directed at some studies (10, 17) that employed these methods in deriving structure–activity relationships. Misleading relationships have been shown to result due to the improper choice of compounds in the data set and to inappropriate structural representation. The approach taken in this article attempts to overcome these problems and, with proper researcher–computer interaction, should provide a simple and consistent means of deriving structure–activity relationships.

In a nonvisual approach to establish the structure–activity relationships, representation of molecular structures in numerical terms is necessary. Coding of molecular structures is thus a key step. Most biological effects may be readily measured and coded, and compounds possessing similar activity are readily identified. However, coding of organic structural groups or features is not as simple. In previous reports, augmented atom fragment (8), substructural fragment (18), and heteropath (9) representations were adopted in molecular structural coding. Several disadvantages are inherent in such procedures, however, especially in the identification of structure–activity relationships: (*a*) molecular structural information is lost by considering fragments of molecules independently of each other, (*b*) several such molecular discriminators may have to be specified in the representation of a given compound, (*c*) redundant codings may have to be included to distinguish between different compounds satisfactorily, and (*d*) interpretation of results in terms of structural prototypes is extremely difficult.

One method associated with pattern recognition, factor analysis, is readily applicable to structure–activity studies (19). Weiner and Weiner (19) used the method directly in deriving structure–activity relationships and obtained readily interpretable results. However, factor analysis is also adaptable as the preprocessing step of a pattern recognition procedure, and this approach is taken in this paper.

Briefly, the method requires that one view molecular structures in a manner consistent with the way medicinal chemists have done in proposing candidate compounds. To retain molecular bonding information, all molecular structures of the data set are superposed upon each other *in a chemically consistent manner.* Major discriminating features between the molecules are then identified. These molecular discriminators are coded so as to distinguish between bioisosteric atoms or groups (20). Following such representation, a factor analytical procedure is applied to reduce the dimensionality of the data matrix. Only a few molecular features need to be specified to classify correctly the compounds of the data set into the pharmacologically distinct categories present.



Scheme I

BASIS OF APPROACH

The basic operations of pattern recognition as applied in this work may be represented as shown in Scheme I. Coding of molecular structures comprising the data set leads to their numerical representation in terms of a discrete number of variables (the input). The input data are then operated on, or preprocessed, to reduce the number of variables required to represent the data, *i.e.*, to reduce the dimensionality. The preprocessing method adopted in this work is principal component analysis, which is but one of the several factor analytical methods available (21–25).

Classification of the output from the preprocessing step is performed next. This step may be done by mathematical techniques such as cluster analysis (26, 27) and discriminant analysis (28) or with the aid of "learning machines" (13, 29–31). In this paper, however, classification is done graphically to provide a clearer conceptual understanding and to point out the parallelism that exists between the classical and pattern recognition approaches for deriving structure–activity relationships. Several detailed accounts in the literature describe the theory and application of various factor analysis techniques (22–25), and only a brief discussion is presented in this article.

A set of M molecules comprising a given data set is coded so as to be represented by N molecular groups or features, giving rise to an $M \times N$ data matrix. This matrix is then used to construct an $N \times N$ correlation matrix. This correlation matrix is next operated on by a factor analytical procedure to extract factors, *i.e.*, to reduce the dimensionality. The two most commonly used methods are principal component analysis and common factor analysis. In principal component analysis (22, 23), one seeks to transform the variables of the data set into a linearly independent set of component variables consisting of the original variables. When N molecular features are chosen as independent variables, then N linear combinations are needed to account for all of the variance in the data. However, in some cases, fewer linear combinations (in comparison with the number of variables in the data matrix) suffice to account for a large proportion of the variance within the data. These may then be used to represent the data and, since these linear combinations are defined to be orthogonal to one another, a reduction in dimensionality is achieved.

In common factor analysis, however, the assumption is made that the variables in a data matrix may be influenced to varying degrees by properties common to *all* variables (*common* variance) and by properties unique to only a few variables (*unique* variance). In other words, every variable selected to represent a specific molecular feature is regarded as reflecting a common set of physical attributes for the majority of compounds in the data set and a unique set of physical attributes for others. Identification is sought of those variables that reflect common physical attributes for the majority of compounds in the data set; all other compounds are treated as unique. The method thus holds the promise of providing some insight regarding the mechanisms leading to major and minor pharmacological actions of drugs. Although some advances have been made in the application of common factor analysis to physiological studies such as olfaction (32), principal component analysis is more readily adaptable in the investigation of structure–activity relationships and is the method used here.

THEORY

The basis for a principal component analysis may be placed in perspective by considering an attempt to transform a multiple regression into a simple linear one. In a multiple regression model, one seeks to relate the biological activities, A, for M compounds to N variables, X, in the

Table I—Pharmacological Classification and Therapeutic Use of Compounds Studied

Compound	Drug	Pharmacological Classification	Therapeutic Use
I	Levarterenol	α-Agonist	Antihypotensive, vasoconstrictor
II	Methoxamine	α-Agonist	Antiarrhythmic, antihypotensive
III	Metaraminol	α-Agonist	Antihypotensive
IV	Ethylnorepinephrine	β-Agonist	Bronchodilator
V	Phenylpropanolamine	α-Agonist, CNS stimulant	Nasal decongestant, anorexiant
VI	Amphetamine	CNS stimulant	Antidepressant, minimal brain dysfunction, anorexiant, narcolepsy
VII	Chlorphentermine	CNS stimulant	Anorexiant
VIII	Phentermine	CNS stimulant	Anorexiant
IX	Hydroxyamphetamine	α-Agonist	Antihypotensive
X	Methyldopa	CNS agent	Antihypertensive
XI	Levodopa	CNS agent	Antiparkinsonian
XII	Epinephrine	α,β-Agonist	Antihypotensive, bronchodilator, decongestant, antiarrhythmic
XIII	Methoxyphenamine	β-Agonist	Bronchodilator
XIV	Phenylephrine	α-Agonist	Decongestant, antihypotensive
XV	Ephedrine	α-Agonist, CNS stimulant	Antihypotensive, decongestant, bronchodilator, antiarrhythmic
XVI	Methamphetamine	CNS stimulant	Anorexiant, minimal brain dysfunction, antidepressant, narcolepsy
XVII	Mephentermine	CNS stimulant, α-agonist	Antihypotensive
XVIII	Propylhexedrine	α-Agonist	Nasal decongestant
XIX	Phenmetrazine	CNS stimulant	Anorexiant
XX	Phendimetrazine	CNS stimulant	Anorexiant
XXI	Isoproterenol	β-Agonist	Antiarrhythmic, bronchodilator
XXII	Isoxsuprine	β-Agonist	Peripheral vascular disease
XXIII	Nylidrin	β-Agonist	Peripheral vascular disease
XXIV	Propranolol	β-Antagonist	Antiarrhythmic, antihypertensive
XXV	Benzphetamine	CNS stimulant	Anorexiant
XXVI	Diethylpropion	CNS stimulant	Anorexiant
XXVII	Phenoxybenzamine	α-Antagonist	Peripheral vascular disease, antihypertensive
XXVIII	Acetylcholine	Parasympathomimetic	Miotic
XXIX	Succinylcholine	Myoneural blocker, parasympathomimetic	Muscle relaxant
XXX	Methacholine	Parasympathomimetic	Glaucoma
XXXI	Carbachol	Parasympathomimetic	Glaucoma, miotic
XXXII	Bethanechol	Parasympathomimetic	Intestinal relaxant
XXXIII	Pilocarpine	Parasympathomimetic	Glaucoma
XXXIV	Ambenonium	Anticholinesterase	Myasthenia gravis
XXXV	Edrophonium	Anticholinesterase	Myasthenia gravis
XXXVI	Physostigmine	Anticholinesterase	Glaucoma, myasthenia gravis
XXXVII	Neostigmine	Anticholinesterase	Myasthenia gravis
XXXVIII	Demecarium	Anticholinesterase	Glaucoma
XXXIX	Pyridostigmine	Anticholinesterase	Myasthenia gravis

form of a linear combination:

$$Ai = b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + \ldots + b_NX_{Ni} + \mu \qquad i = 1, 2, \ldots, M \quad \text{(Eq. 1)}$$

However, in evaluating the least-squares estimates of the b coefficients, the values of the coefficients are necessarily a function of the covariance between the independent variables, X, as well as of the covariance between the biological activities and the independent variables. One may thus seek to account for the covariance between X prior to attempting a regression analysis.

In making this attempt, a transformed variable, T, may be defined as a linear combination of the independent variables, X:

$$Ti = m_1X_{1i} + m_2X_{2i} + m_3X_{3i} + \ldots + m_NX_{Ni} \qquad i = 1, 2, \ldots, M \quad \text{(Eq. 2)}$$

This equation is fundamental to all factor analysis procedures (principal component analysis in this case). In the simple case where only three independent variables are involved, Eq. 2 is written as:

$$Ti = m_1X_{1i} + m_2X_{2i} + m_3X_{3i} \qquad i = 1, 2, \ldots, M \quad \text{(Eq. 3)}$$

Squaring both sides of Eq. 3 yields:

$$\sum_i Ti^2 = \sum_i (m_1X_{1i} + m_2X_{2i} + m_3X_{3i})^2 \quad \text{(Eq. 4)}$$

where $\sum_i$ indicates that the summation is over all compounds (1 to M) unless otherwise specified. For convenience of mathematical manipulation, u is defined by the expression:

$$u = \sum_i Ti^2 \quad \text{(Eq. 5)}$$

so that:

$$u = \sum_i (m_1X_{1i} + m_2X_{2i} + m_3X_{3i})^2 \quad \text{(Eq. 6)}$$

The coefficients m_1, m_2, and m_3 may be defined to be orthogonal; *i.e.*, they are subject to the conditional equation:

$$v = m_1^2 + m_2^2 + m_3^2 - 1 = 0 \quad \text{(Eq. 7)}$$

(v is introduced for mathematical convenience). Minimization of the coefficients of Eq. 6 under the restraint imposed by Eq. 7 can be accomplished with the aid of Lagrangian multipliers (33). Differentiation with respect to each coefficient thus leads to:

$$\frac{\partial u}{\partial m_1} - \lambda \frac{\partial v}{\partial m_1} = 0 \quad \text{(Eq. 8a)}$$

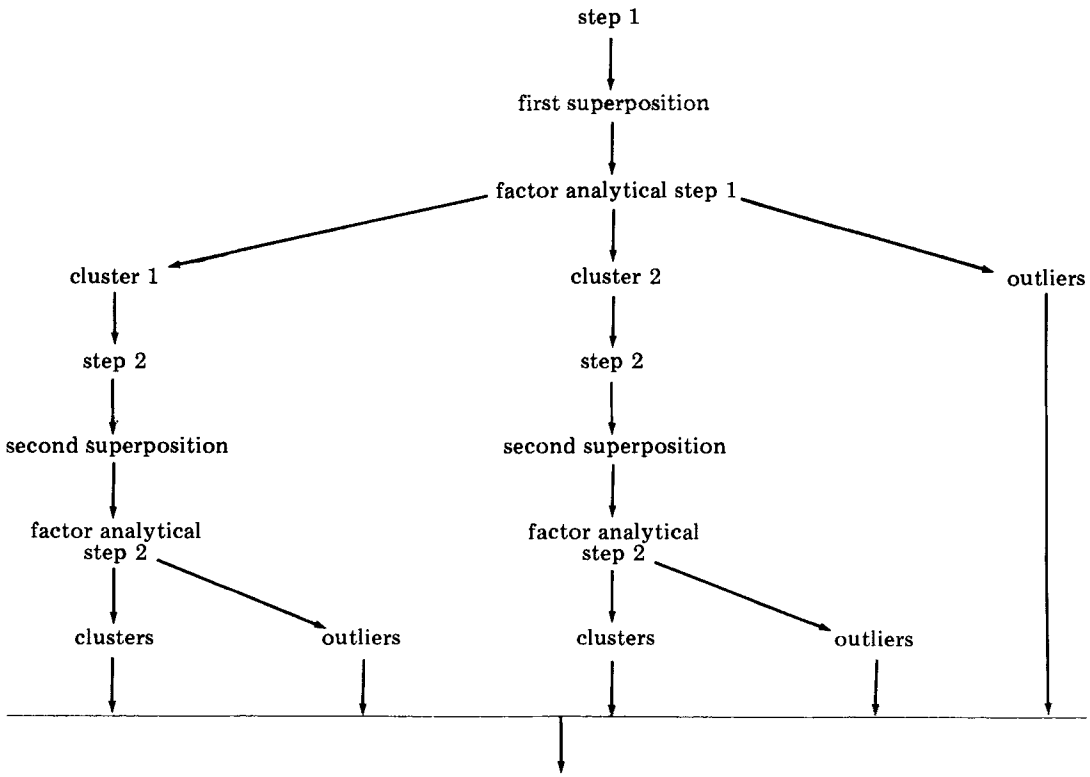
$$\frac{\partial u}{\partial m_2} - \lambda \frac{\partial v}{\partial m_2} = 0 \quad \text{(Eq. 8b)}$$

$$\frac{\partial u}{\partial m_3} - \lambda \frac{v}{\partial m_3} = 0 \quad \text{(Eq. 8c)}$$

Evaluation of Eqs. 8a–8c provides the set of relationships that may be written in matrix formalism as:

$$\begin{vmatrix} \sum X_1^2 - \lambda & \sum X_1X_2 & \sum X_1X_3 \\ \sum X_2X_1 & \sum X_2^2 - \lambda & \sum X_2X_3 \\ \sum X_3X_1 & \sum X_3X_2 & \sum X_3^2 - \lambda \end{vmatrix} \begin{vmatrix} m_1 \\ m_2 \\ m_3 \end{vmatrix} = 0 \quad \text{(Eq. 9)}$$

In Eq. 9, the matrix involving the independent variables is a covariance matrix. However, a technical problem arises in such cases as a consequence of the relative magnitudes of the values used to represent the



Scheme II

independent variables. Variables that are numerically greater in a data matrix will lead to covariances of a numerically large magnitude; therefore, these variables will be designated, artificially, as the more important contributors in a given linear combination. The potential for an artifact of this type in Fujita–Iwasa–Hansch (34) analyses has been pointed out (35). Autoscaling of variables is a means of avoiding this problem (13). Autoscaling involves the use of correlation coefficients in the place of covariances when establishing normal equations such as is represented by Eq. 9. The elements of a correlation matrix are defined by the relationship:

$$r_{rs} = \frac{\sum_i (X_{ri} - \overline{X}_r)(X_{si} - \overline{X}_s)}{\sqrt{\sum_i (X_{ri} - \overline{X}_r)^2 \sum_i (X_{si} - \overline{X}_s)^2}} \qquad \text{(Eq. 10)}$$

where r and s designate the variables whose intercorrelation is being evaluated. Equation 9 may thus be rewritten as:

$$\begin{vmatrix} r_{11} - \lambda & r_{12} & r_{13} \\ r_{21} & r_{22} - \lambda & r_{23} \\ r_{31} & r_{32} & r_{33} - \lambda \end{vmatrix} \begin{vmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{vmatrix} = 0 \qquad \text{(Eq. 11)}$$

The β coefficients of Eq. 11 are related to the m coefficients of Eq. 9 by the expression:

$$m_k = \beta_k \lambda^{1/2} \qquad \text{(Eq. 12)}$$

By using correlation coefficients, any one of the elements of the correlation matrix can vary only between ±1 and 0, so no element of the matrix is overweighted.

The solutions for Eq. 11 can be obtained by first determining the eigenvalues, λ_i, which characterize the correlation matrix. As many eigenvalues will be obtained as there are variables in the data matrix. For each eigenvalue, an associated set of eigenvectors, T_i, is obtained.

A relative measure of the information content of the linear combinations T_i is provided by the magnitudes of the associated eigenvalues. The total variance explained by one or more of the eigenvalues is given by the expression:

$$\text{fraction "explained" variance} = \frac{\sum_{i=1}^{L} \lambda i}{N} \qquad \text{(Eq. 13)}$$

where L is the number of eigenvalues selected, and N is the number of variables in the data matrix. It is apparent that all of the eigenvalues taken together represent the data exactly; *i.e.*, the fraction of explained variance is 1. However, in some cases it may be sufficient to choose only a few high value eigenvectors so that only a predetermined fraction of the variance in the data is explained. Thus, one may choose those high value eigenvectors that explain greater than 0.95 of the data, as is often the case in physicochemical applications, or one may choose to work with an approximated form of the data matrix, selecting only those eigenvectors whose associated eigenvalues are equal to 1 or greater (23). The latter approach is taken in this paper, because fewer numbers of eigenvectors are required to represent the data graphically and because such an approximation is quite adequate for the present purposes.

Principal component analysis presents a means by which optimization procedures involving regression methods such as Fujita–Iwasa–Hansch analyses (34) can be made less prone to error due to intercorrelation between variables. Correlation analysis utilizing principal component analysis as a preprocessing method would thus be based on the multiple regression model in the form:

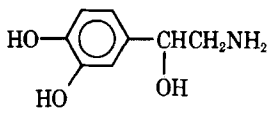
$$Ai = \sum_{t=1}^{K} C_t T_{ti} + \mu \qquad i = 1, 2, \ldots, M \qquad \text{(Eq. 14)}$$

where the K orthogonal variables, T, explain the major proportion of the variance in the data. An interface thus exists between pattern recognition techniques and the more widely used quantitative structure–activity methods.

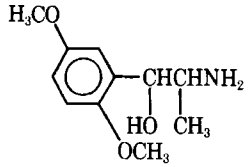
METHOD

The 39 drugs comprising the data set are listed in Table I along with their known pharmacological classifications (36) and therapeutic uses (37). They consist of various pharmacological types, including α- and β-adrenergic agents, cholinergic agents, and central nervous system (CNS) stimulants. The study was directed toward separating these compounds into the distinct pharmacological groups they are known to fall into and to separate out the unrelated compounds.

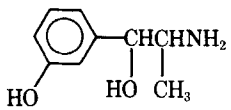
Scheme II is an example of a two-step pattern recognition procedure as followed in this work. In the first step, a preliminary factor analysis is performed to separate broadly the entire data set. Classification is then accomplished by plotting the output from the preprocessing step. Clusters



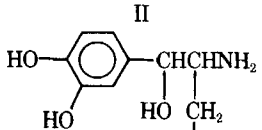
I



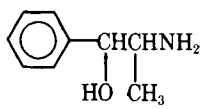
II



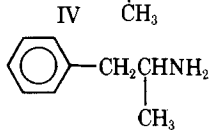
III



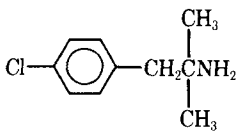
IV



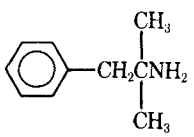
V



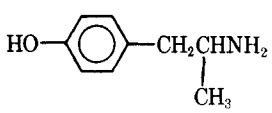
VI



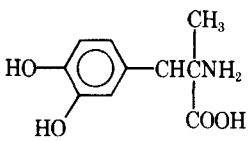
VII



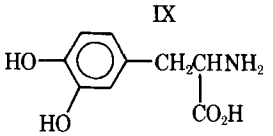
VIII



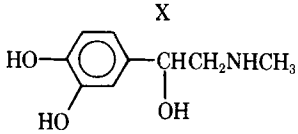
IX



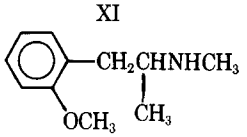
X



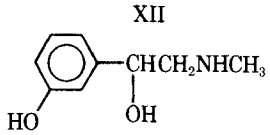
XI



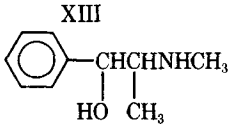
XII



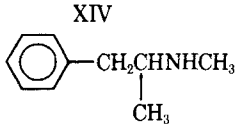
XIII



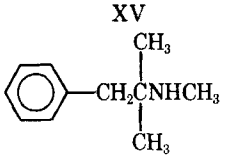
XIV



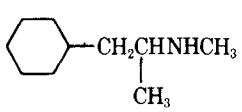
XV



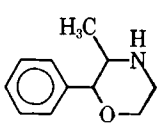
XVI



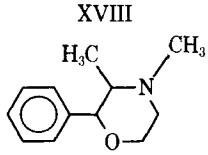
XVII



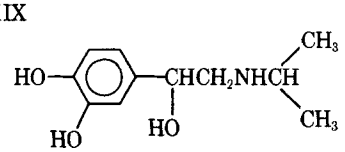
XVIII



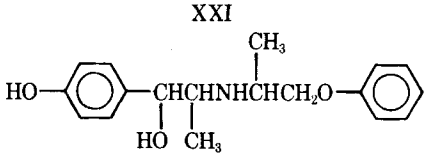
XIX



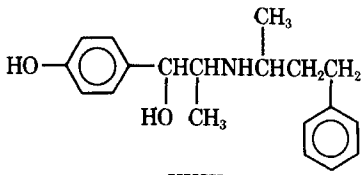
XX



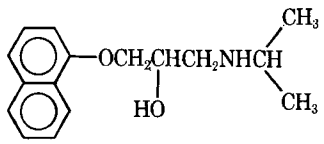
XXI



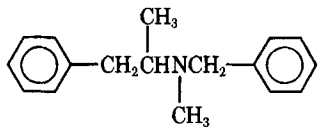
XXII



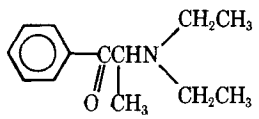
XXIII



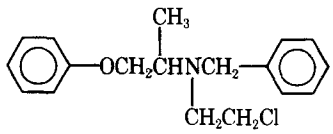
XXIV



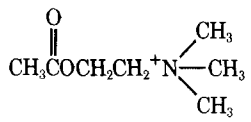
XXV



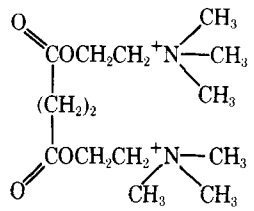
XXVI



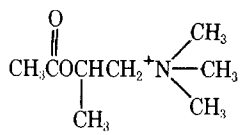
XXVII



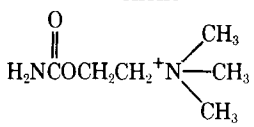
XXVIII



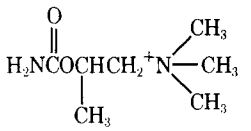
XXIX



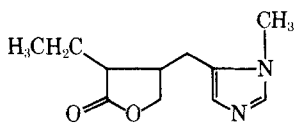
XXX



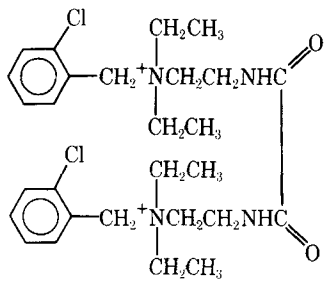
XXXI



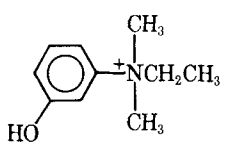
XXXII



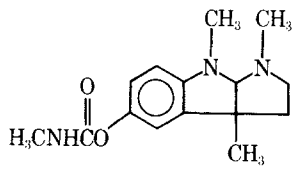
XXXIII



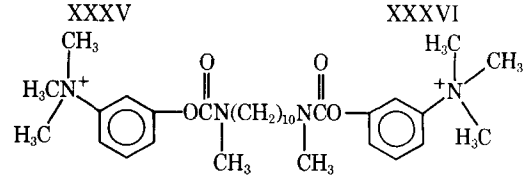
XXXIV



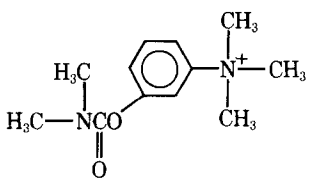
XXXV



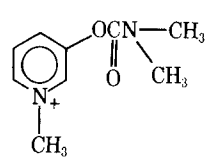
XXXVI



XXXVIII



XXXVII



XXXIX

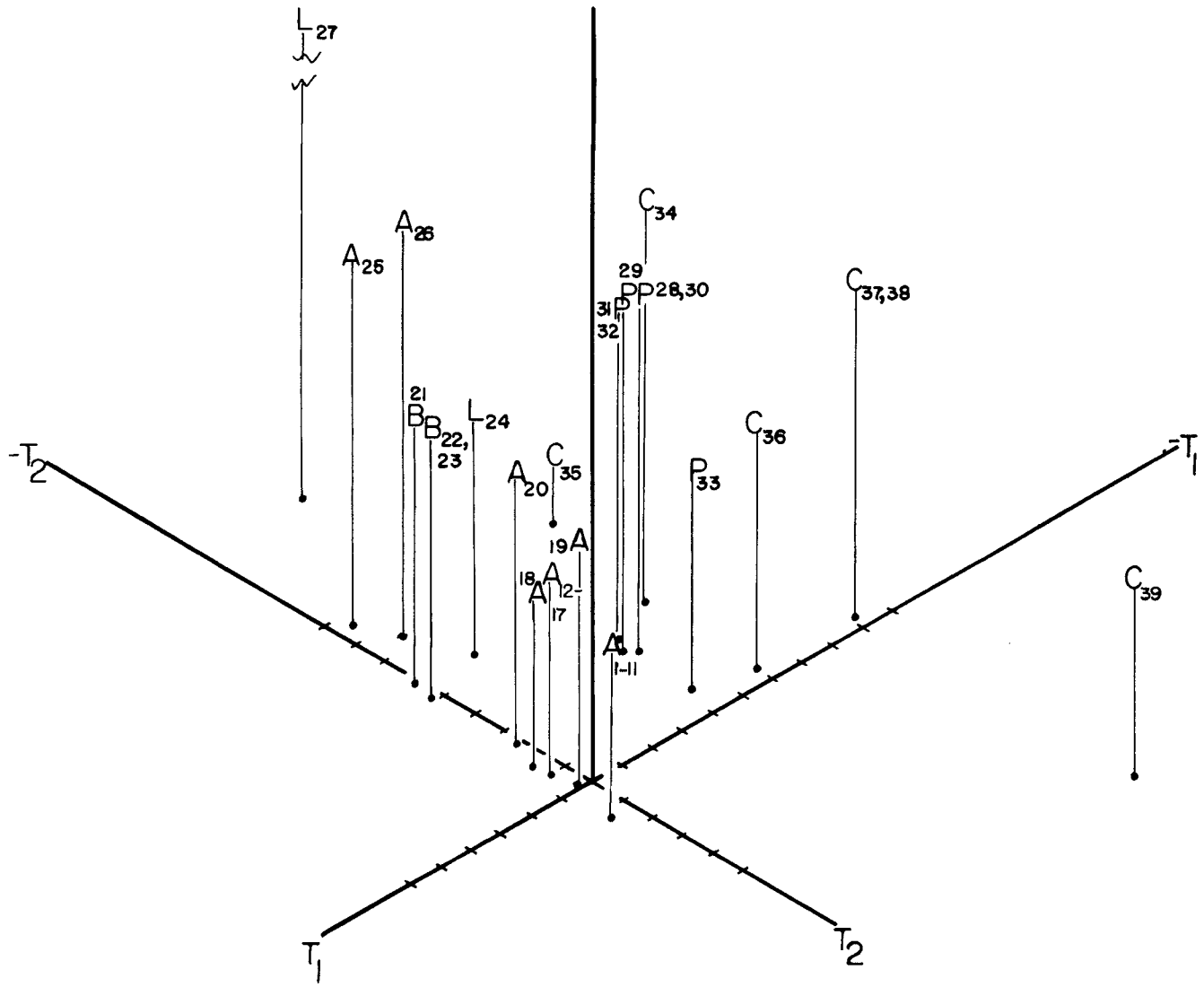


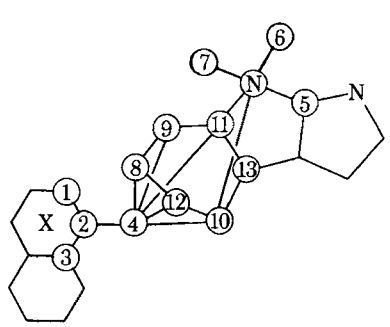
Figure 1—*Three-dimensional plot of factor coordinates from factor analytical step 1. A = α-adrenergic or CNS agent, B = β-adrenergic agent, C = anticholinesterase, L = adrenergic blocker, and P = parasympathomimetic. (Arabic numbers correspond to the roman numerals for the compounds used in the text.)*

so obtained are then factor analyzed separately to yield smaller clusters. Such a sequential procedure has two distinct advantages: (*a*) only a few discriminating features need to be specified in each step, and (*b*) since only a few discriminators are included in the data matrix, easily interpretable results are obtained, the solutions being in two or three dimensions.

Step 1—*Superpositioning of Structures*—Since only variant molecular features are included in the basic equation (Eq. 2), all of the compounds of the data set must first be superimposed to generate the simplest possible superposed structure or "superstructure." This composite structure is so designed that all molecules of the data set can be superposed on it in a consistent manner. This is a key step of the procedure adopted here, since the complexity of the final results depends on the complexity of the superposed structure chosen. This structure has no molecular significance; it is only a geometrical construction. It may be described as a hypothetical parent structure for all compounds of the data set.

In the present case, examination of the structures of I–XXXIX shows a wide variation of structural types. The task of superimposition of all these structures is considerably simplified if representative structures are selected from among structurally related compounds. In the given data set, one might pick XXI, XXIV, XXVII, XXXIII, and XXXVI as being representatives. All of these structures have the common feature X—$(C)_n$—N, where X may be a carboxylic acid derivative or aromatic ring, *n* is the number of carbon atoms intervening, and the amine nitrogen N may be variously substituted. The superposition of the molecules is achieved by superposing the X groups and the N groups while, at the same time, carefully accounting for the length of the carbon chain between them (XL).

In the case of phenethylamines like isoproterenol (XXI) and amphetamine (VI), the aromatic ring is superposed on ring X (see Structure XL), the carbon atom joined to the ethylamine group being placed on position 2. The two-carbon chain is superposed on 4 and 11; the isopropylamino group is on N, 5, and 6; and all other positions (7, 8, 9, 10, 12, 13) are nonexistent. For a molecule like acetylcholine, the carbonyl group is superposable on 2 and 3, the esteratic methyl group on 1, the ether oxygen on 4, the two-carbon chain on 9 and 11, and the trimethylammonium function on N, 5, 6, and 7. Positions 8, 10, 12, and 13 are absent



XL

Table II—Molecule-Feature Matrix

Compound	Position[a] 1	2	3	4	5	6	7	8	9	10	11	12	13
I	C_r	C_r	C_r	C_a	H	H	—	—	—	—	C_a	—	—
II	C_r	C_r	C_r	C_a	H	H	—	—	—	—	C_a	—	—
III	C_r	C_r	C_r	C_a	H	H	—	—	—	—	C_a	—	—
IV	C_r	C_r	C_r	C_a	H	H	—	—	—	—	C_a	—	—
V	C_r	C_r	C_r	C_a	H	H	—	—	—	—	C_a	—	—
VI	C_r	C_r	C_r	C_a	H	H	—	—	—	—	C_a	—	—
VII	C_r	C_r	C_r	C_a	H	H	—	—	—	—	C_a	—	—
VIII	C_r	C_r	C_r	C_a	H	H	—	—	—	—	C_a	—	—
IX	C_r	C_r	C_r	C_a	H	H	—	—	—	—	C_a	—	—
X	C_r	C_r	C_r	C_a	H	H	—	—	—	—	C_a	—	—
XI	C_r	C_r	C_r	C_a	H	H	—	—	—	—	C_a	—	—
XII	C_r	C_r	C_r	C_a	H	CH_3	—	—	—	—	C_a	—	—
XIII	C_r	C_r	C_r	C_a	H	CH_3	—	—	—	—	C_a	—	—
XIV	C_r	C_r	C_r	C_a	H	CH_3	—	—	—	—	C_a	—	—
XV	C_r	C_r	C_r	C_a	H	CH_3	—	—	—	—	C_a	—	—
XVI	C_r	C_r	C_r	C_a	H	CH_3	—	—	—	—	C_a	—	—
XVII	C_r	C_r	C_r	C_a	H	CH_3	—	—	—	—	C_a	—	—
XVIII	C_a	C_a	C_a	C_a	H	CH_3	—	—	—	—	C_a	—	—
XIX	C_r	C_r	C_r	C_a	CH_2	H	—	—	—	—	C_a	—	—
XX	C_r	C_r	C_r	C_a	CH_2	$CH(CH_3)_2$	—	—	—	—	C_a	—	—
XXI	C_r	C_r	C_r	C_a	H	$CH(CH_3)CH(CH_3)_2$	—	—	—	—	C_a	—	—
XXII	C_r	C_r	C_r	C_a	H	$CH(CH_2OC_6H_5)CH_3$	—	—	—	—	C_a	—	—
XXIII	C_r	C_r	C_r	C_a	H	$CH(CH_2CH_2C_6H_5)CH_3$	—	—	—	—	C_a	—	—
XXIV	C_r	C_r	C_r	O	H	$CH(CH_3)_2$	—	C_a	C_a	—	C_a	—	—
XXV	C_r	C_r	C_r	C_a	CH_3	$CH_2C_6H_5$	—	—	—	—	C_a	—	—
XXVI	C_r	C_r	C_r	C_a	CH_2CH_3	CH_2CH_3	—	—	—	—	C_a	—	—
XXVII	C_r	C_r	C_r	O	CH_2CH_2Cl	$CH_2C_6H_5$	—	—	—	—	C_a	—	—
XXVIII	CH_3	C_o	=O	O	CH_3	CH_3	CH_3	—	C_a	—	C_a	—	—
XXIX	CH_2	C_o	=O	O	CH_3	CH_3	CH_3	—	C_a	—	C_a	—	—
XXX	CH_3	C_o	=O	O	CH_3	CH_3	CH_3	—	C_a	—	C_a	—	—
XXXI	NH_2	C_o	=O	O	CH_3	CH_3	CH_3	—	C_a	—	C_a	—	—
XXXII	NH_2	C_o	=O	O	CH_3	CH_3	CH_3	—	C_a	—	C_a	—	—
XXXIII	CH	C_o	=O	O	HC=	CH_3	—	C_a	C_a	—	C_a	C_a	—
XXXIV	CONH	C_o	=O	NH	CH_2	CH_2CH_3	CH_2CH_3	—	C_a	—	C_a	—	—
XXXV	—	H	—	O	CH_3	CH_2CH_3	CH_2CH_3	C_r	C_r	C_r	C_r	C_r	C_r
XXXVI	$NHCH_3$	C_o	=O	O	CH	CH_3	—	C_r	C_r	C_r	C_r	C_r	C_r
XXXVII	$N(CH_3)_2$	C_o	=O	O	CH_3	CH_3	CH_3	C_r	C_r	C_r	C_r	C_r	C_r
XXXVIII	$N(CH_3)_2$	C_o	=O	O	CH_3	CH_3	CH_3	C_r	C_r	C_r	C_r	C_r	C_r
XXXIX	$N(CH_3)_2$	C_o	=O	O	—	—	CH_3	C_r	C_r	C_r	C_r	C_r	—

[a] C_r = aromatic carbon atom, C_a = aliphatic carbon atom, and C_o = carbonyl carbon atom.

in this case. For compounds like neostigmine or physostigmine, the ester or amide group is treated just as for acetylcholine; the aromatic ring is superposed on the six-membered ring formed by positions 8, 9, 11, 13, 10, and 12. If the esteratic group is *para* to the amine function, then the former is linked to the aromatic ring by the 4–12 bond; if the esteratic group is *meta* to the aromatic ring, it is linked to the latter by the 4–10 bond.

Two of the molecules of the data set, pilocarpine (XXXIII) and pyridostigmine (XXXIX), are structurally distinct from the rest; they may be superposed on Structure XL as follows. The ester group of pilocarpine is treated just as for acetylcholine; the four-carbon chain between the ester group and the amine function is superposable on positions 12, 8, 9, and 11. The superpositioning of the nitrogen-containing rings is apparent from the figure. In the case of pyridostigimine, the pyridine ring is superposed on N, 11, 9, 12, and 10. Position 4 (ether oxygen) is linked to 12 to represent the *meta*-substitution.

All other structures are similarly superposable; Table II shows exactly how each compound was treated.

Selection of Discriminators—Discriminators are groups, bonds, or other functionalities that serve to distinguish between compounds of a data set. When a wide variety of structural types are being studied, the number of variables in the data matrix will be inordinately large if *all* structural differences are designated as discriminators. Therefore, it is desirable to specify the minimum number of discriminators required to differentiate between the compounds of the data set.

In the present study, more than one factor analytical step is to be applied (Scheme II). In the first step, only a broad classification is sought, so only major structural differences need be specified while selecting discriminators. Thus, fine distinctions such as aromatic or aliphatic substitution patterns may be ignored while major distinctions such as those between aromatic and aliphatic groups should be specified. On this basis, positions 1–13 of XL may be chosen as discriminators. However, closer examination of Structures I–XXXIX and of Table II shows that positions 11–13 are not major discriminators; they represent only two variations in structure (aromatic or aliphatic carbon atoms). Furthermore, these positions are equally well represented by others, *e.g.*, 12 and 13 by

Table III—Descriptor Code for Designating Discriminator Groups

Group[a]	Atomic/Structural Constant	Code
Cl	5.84	1.00
C_r	3.37	0.58
C_a	2.59	0.44
H	1.03	0.18
CH_3	5.65	0.97
CH_2CH_3	10.30	1.76
CH_2	4.65	0.80
$CH(CH_3)_2$	14.98	2.57
$CHCH_3$	13.95[b]	2.39
CH_2		
$CH_2(C_r)_3$[c]	16.81[b]	2.88
CH	3.62	0.62
NH_2	4.44	0.76
$NHCH_3$	9.06[b]	1.55
$N(CH_3)_2$	13.68[b]	2.34
CONH	8.01[b]	1.37
C_o	0.79[b]	0.13
=O	3.81[b]	0.65
O	1.76	0.30
NH	3.62	0.62
=CH	1.82[b]	0.31
CH_2CH_2Cl	15.14[b]	2.59
OH_a[d]	2.55	0.44
OH_p[e]	3.55	0.60
OCH_3	7.41	1.26
COOH	7.23	1.24

[a] C_r = aromatic carbon, C_o = carbonyl carbon, and C_a = aliphatic carbon. [b] Value was calculated. [c] Only the aromatic carbon attached to the methylene group and the two *ortho*-carbon atoms were included to be consistent in designating discriminators; the value of 16.81 was calculated using the molar refraction value for the benzene ring (38). [d] Aliphatic hydroxyl group. [e] Phenolic hydroxyl group.

8 for all compounds [except for pilocarpine (XXXIII) and pyridostigmine (XXXIX), which are in any case different from the other compounds of the data set]. Similarly, position 11 is represented by position 4 in I–XXVII and by position 9 in XXVIII–XXXIX. Positions 11–13 may therefore be ignored.

Designation of Discriminators: Coding—The discriminators selected are designated in numerical terms by a descriptor code. One has to choose a property of the groups that can adequately represent and distinguish between them. It is seen that many of the substituents involved are bioisosteric, *e.g.*, $COOCH_3$, $CONH_2$, and $CON(CH_3)_2$. Since it is generally true that classical isosteric groups have similar molar refractivities, molar refraction values were used in the coding (38). The molar refraction value of chlorine (5.84) was selected as standard, and all other codes were determined relatively. The molar refraction values used and the descriptor codes adopted are given in Table III. Several molar refraction values were calculated by using the additive-constitutive property.

With the codes given in Table III, the molecule-discriminator data matrix shown in Table IV was constructed. Comparison of Structures I–XXXIX with XL and of Tables II–IV will indicate exactly how each compound of the data set was numerically designated.

Preprocessing: Factor Analytical Step 1—By using the coded molecule-discriminator data matrix of Table IV, the correlation matrix shown in Table V was constructed (as described under *Theory*). The solution of Eq. 13 gave three eigenvalues whose values were greater than unity. These values, together with the associated eigenvectors, are given in Table VI. The three eigenvalues chosen explain $(4.87 + 1.61 + 1.32)/10 = 0.78$ of the variance. This value is not very high, but only major classification patterns are sought in this part of the analysis. That an explained variance of 78% is sufficient for the present purposes is obvious upon examination of the clustering pattern obtained (Fig. 1).

Classification: Plotting—By using the eigenvectors associated with the three eigenvalues, three factor coordinates, t_1, t_2, and t_3, were calculated for each compound of the data set. The points were then plotted by simple solid geometrical construction (Fig. 1). After plotting each point, its associated biological activity was used to label it in the distri-

Table IV—Coded Molecule-Feature Data Matrix

Compound	Position									
	1	2	3	4	5	6	7	8	9	10
I	0.58	0.58	0.58	0.44	0.18	0.18	0.00	0.00	0.00	0.00
II	0.58	0.58	0.58	0.44	0.18	0.18	0.00	0.00	0.00	0.00
III	0.58	0.58	0.58	0.44	0.18	0.18	0.00	0.00	0.00	0.00
IV	0.58	0.58	0.58	0.44	0.18	0.18	0.00	0.00	0.00	0.00
V	0.58	0.58	0.58	0.44	0.18	0.18	0.00	0.00	0.00	0.00
VI	0.58	0.58	0.58	0.44	0.18	0.18	0.00	0.00	0.00	0.00
VII	0.58	0.58	0.58	0.44	0.18	0.18	0.00	0.00	0.00	0.00
VIII	0.58	0.58	0.58	0.44	0.18	0.18	0.00	0.00	0.00	0.00
IX	0.58	0.58	0.58	0.44	0.18	0.18	0.00	0.00	0.00	0.00
X	0.58	0.58	0.58	0.44	0.18	0.18	0.00	0.00	0.00	0.00
XI	0.58	0.58	0.58	0.44	0.18	0.18	0.00	0.00	0.00	0.00
XII	0.58	0.58	0.58	0.44	0.18	0.97	0.00	0.00	0.00	0.00
XIII	0.58	0.58	0.58	0.44	0.18	0.97	0.00	0.00	0.00	0.00
XIV	0.58	0.58	0.58	0.44	0.18	0.97	0.00	0.00	0.00	0.00
XV	0.58	0.58	0.58	0.44	0.18	0.97	0.00	0.00	0.00	0.00
XVI	0.58	0.58	0.58	0.44	0.18	0.97	0.00	0.00	0.00	0.00
XVII	0.58	0.58	0.58	0.44	0.18	0.97	0.00	0.00	0.00	0.00
XVIII	0.44	0.44	0.44	0.44	0.18	0.97	0.00	0.00	0.00	0.00
XIX	0.58	0.58	0.58	0.44	0.80	0.18	0.00	0.00	0.00	0.00
XX	0.58	0.58	0.58	0.44	0.80	0.97	0.00	0.00	0.00	0.00
XXI	0.58	0.58	0.58	0.44	0.18	2.57	0.00	0.00	0.00	0.00
XXII	0.58	0.58	0.58	0.44	0.18	2.39	0.00	0.00	0.00	0.00
XXIII	0.58	0.58	0.58	0.44	0.18	2.39	0.00	0.00	0.00	0.00
XXIV	0.58	0.58	0.58	0.30	0.18	2.57	0.00	0.44	0.44	0.00
XXV	0.58	0.58	0.58	0.44	0.97	2.88	0.00	0.00	0.00	0.00
XXVI	0.58	0.58	0.58	0.44	1.76	1.76	0.00	0.00	0.00	0.00
XXVII	0.58	0.58	0.58	0.30	2.59	2.88	0.00	0.00	0.44	0.00
XXVIII	0.97	0.13	0.65	0.30	0.97	0.97	0.97	0.00	0.44	0.00
XXIX	0.80	0.13	0.65	0.30	0.97	0.97	0.97	0.00	0.44	0.00
XXX	0.97	0.13	0.65	0.30	0.97	0.97	0.97	0.00	0.44	0.00
XXXI	0.76	0.13	0.65	0.30	0.97	0.97	0.97	0.00	0.44	0.00
XXXII	0.76	0.13	0.65	0.30	0.97	0.97	0.97	0.00	0.44	0.00
XXXIII	0.62	0.13	0.65	0.30	0.31	0.97	0.00	0.44	0.44	0.00
XXXIV	1.37	0.13	0.65	0.62	0.80	1.76	1.76	0.58	0.58	0.58
XXXV	0.00	0.18	0.00	0.30	0.97	1.76	1.76	0.58	0.58	0.58
XXXVI	1.55	0.13	0.65	0.30	0.62	0.97	0.00	0.58	0.58	0.58
XXXVII	2.34	0.13	0.65	0.30	0.97	0.97	0.97	0.58	0.58	0.58
XXXVIII	2.34	0.13	0.65	0.30	0.97	0.97	0.97	0.58	0.58	0.58
XXXIX	2.34	0.13	0.65	0.30	0.00	0.00	0.97	0.58	0.58	0.58

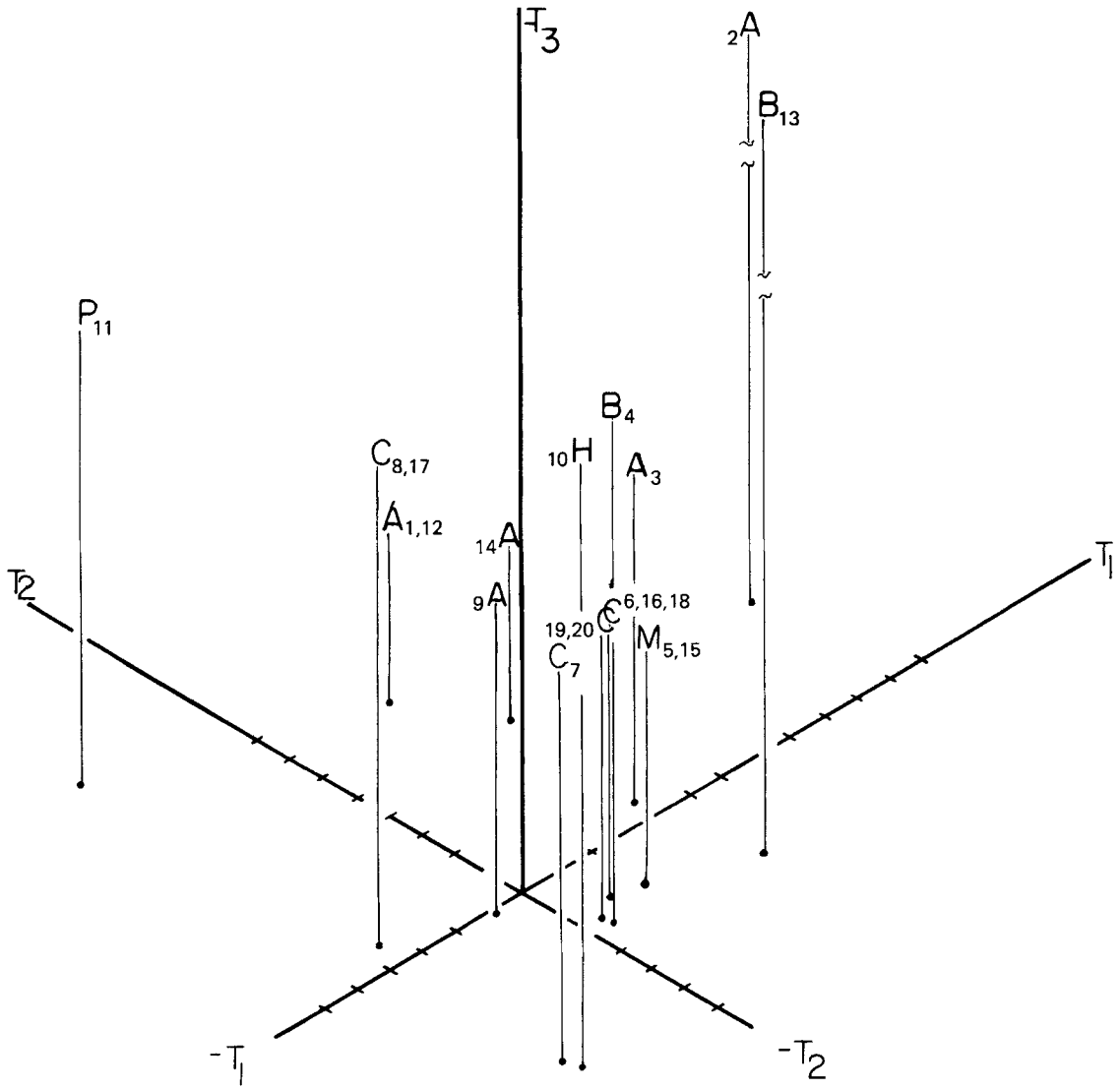


Figure 2—*Three-dimensional plot of factor coordinates from factor analytical step 2. A = α-adrenergic agonist, B = β-adrenergic agonist, C = CNS stimulant, H = antihypertensive, M = mixed-acting agent (α-adrenergic agonist and CNS stimulant), and P = antiparkinsonian. (Arabic numbers correspond to the roman numerals for the compounds used in the text.)*

bution (B for β-adrenergic agent, P for parasympathomimetic, C for anticholinesterase, A for α-adrenergic or CNS agent, and L for adrenergic blocker). Examination of Fig. 1 shows distinct patterns in the distribution. Adrenergics and cholinergics are well separated, and the β-agonists form a well-resolved cluster. The parasympathomimetics form a distinct cluster compared to the scattered distribution of the anticholinesterases. Phenoxybenzamine (XXVII), the only α-blocker, is far removed from the rest of the data set; propranolol (XXIV), the only β-blocker, is removed from, but close to, the β-agonists.

Pilocarpine (XXXIII) appears to occupy an ambiguous position; however, this compound is structurally and pharmacologically quite different from the rest so it can be regarded as well resolved. Only two compounds, ethylnorepinephrine (IV) and methoxyphenamine (XIII), are not properly classified. They are found within the cluster of α-adrenergic and CNS agents (these agents are primarily β-adrenergic agonists). However, they are well resolved after the second factor analytical step, when fine substitution patterns are considered.

Step 2—The distribution pattern of Fig. 1 indicates that, although the major pharmacological classes have been correctly identified and well separated, further separation is desirable among the adrenergic and CNS agents. The group of 20 compounds clustering around the T_3 axis (I–XX, Fig. 1) is fairly large and is subjected to another analysis. The list of

Table V—Correlation Matrix for Factor Analytical Step 1

Position	Position									
	1	2	3	4	5	6	7	8	9	10
1	1.000	−0.605	0.460	−0.352	0.130	−0.115	0.445	0.616	0.611	0.705
2		1.000	−0.107	0.626	−0.329	0.002	−0.822	−0.564	−0.891	−0.560
3			1.000	−0.007	−0.008	−0.143	−0.178	−0.177	0.034	−0.229
4				1.000	−0.394	−0.103	−0.358	−0.607	−0.763	−0.502
5					1.000	0.445	0.359	0.067	0.472	0.134
6						1.000	0.079	0.063	0.191	−0.037
7							1.000	0.382	0.754	0.477
8								1.000	0.722	0.886
9									1.000	0.649
10										1.000

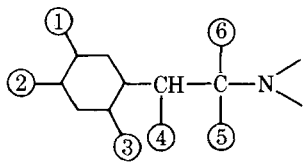
Table VI—Eigenvalues and Associated Eigenvectors for Step 1

	Eigenvalue I = 4.87	Eigenvalue II = 1.61	Eigenvalue III = 1.32
β_1	−0.33[a]	0.39	0.22
β_2	0.40	−0.03	−0.14
β_3	0.00	0.38	0.74
β_4	0.34	0.10	−0.04
β_5	−0.19	−0.51	0.35
β_6	−0.05	−0.60	0.12
β_7	−0.34	−0.13	0.00
β_8	−0.37	0.13	−0.34
β_9	−0.43	−0.09	0.09
β_{10}	−0.37	0.16	−0.35

[a] Values are the associated eigenvector coefficients.

compounds is given in Table VII, which also includes the molecule-feature matrix. The distinction in this case (compared to Step 1, Table III) is that fine substitution patterns are taken into account.

By following the same operational procedures outlined earlier, a composite superstructure for all of the molecules of the data set was generated (XLI). Structure XLI is considerably simplified from Structure XL, since only a limited data set is involved and preliminary separation has already been effected.



XLI

By using the molecule-feature data matrix shown in Table VII, the coded molecule-discriminator matrix was constructed as explained earlier, using the coding given in Table III. Upon factor analyzing the data matrix, three eigenvalues with magnitudes greater than one were obtained; these values, together with their associated eigenvectors, are given in Table VIII. The three eigenvalues together explain 75% of the variance.

The factor coordinates of each compound were obtained as described earlier, and the three-dimensional plot is shown in Fig. 2. The adrenergic agents are well separated from the CNS agents. The two β-adrenergic agents present, ethylnorepinephrine (IV) and methoxyphenamine (XIII), are well resolved. The mixed acting agents, ephedrine (XV) and phenylpropanolamine (V), occupy a position between the α-adrenergic and CNS stimulant agents. This is also characteristic of their pharmacological action, because they have both types of activity. Methyldopa (X) and levodopa (XI) occupy distinct positions, reflecting their characteristically different pharmacological activities.

Table VII—Molecule-Feature Matrix for Factor Analytical Step 2

Compound	Position 1	2	3	4	5	6
I	OH_p[a]	OH_p	H	OH_l[b]	H	H
II	OCH_3	H	OCH_3	OH_l	H	H
III	OH_p	H	H	OH_l	H	CH_3
IV	OH_p	OH_p	H	OH_l	H	CH_2 CH_3
V	H	H	H	OH_l	H	CH_3
VI	H	H	H	H	H	CH_3
VII	H	Cl	H	H	CH_3	CH_3
VIII	H	H	H	H	CH_3	CH_3
IX	H	OH_p	H	H	H	CH_3
X	OH_p	OH_p	H	H	COOH	CH_3
XI	OH_p	OH_p	H	H	COOH	H
XII	OH_p	OH_p	H	OH_l	H	H
XIII	H	H	OCH_3	H	H	CH_3
XIV	OH_p	H	H	OH_l	H	H
XV	H	H	H	OH_l	H	CH_3
XVI	H	H	H	H	H	CH_3
XVII	H	H	H	H	CH_3	CH_3
XVIII	H	H	H	H	H	CH_3
XIX	H	H	H	O	H	CH_3
XX	H	H	H	O	H	CH_3

[a] Phenolic hydroxyl. [b] Aliphatic hydroxyl.

Table VIII—Eigenvalues and Associated Eigenvectors for Step 2

	Eigenvalue I = 1.96	Eigenvalue II = 1.54	Eigenvalue III = 1.10
β_1	0.43[a]	0.52	−0.30
β_2	−0.29	0.55	0.00
β_3	0.37	−0.06	−0.76
β_4	0.57	0.18	0.38
β_5	−0.51	0.35	−0.32
β_6	−0.06	−0.51	−0.30

[a] Values are the associated eigenvector coefficients.

CONCLUSIONS

When dealing with a large number of structurally diverse compounds, a likely problem with the method adopted here is incorrect structural superposition. Attempts to overcome this difficulty are underway.

To researchers accustomed to setting a cutoff level of 0.95 or greater for experimental accuracy, the approximation of 0.75 may not be satisfying. However, a highly accurate reproduction of the data is not essential here; only an approximation of the data is sought so that the major distinctions between the elements of the data set can readily be identified. Furthermore, graphical representation of the results, which is possible by such an approach, is directly interpretable and is conceptually simpler than the mathematical methods available.

To summarize, a simple factor analytical approach was successfully applied to classify a set of 39 therapeutic agents using organic structural information. The major pharmacological groups present were correctly identified and separated. Pharmacologically unrelated compounds such as levodopa (an antiparkinsonian) and methyldopa (an antihypertensive) were separated out. Inspection of the clusters indicated some degree of order; for example, ephedrine and phenylpropanolamine occupied the border region between α-adrenergic agents and CNS stimulant agents. The results show that factor analysis can be employed advantageously as the preprocessing step of a pattern recognition scheme in the investigation of structure–activity relationships.

REFERENCES

(1) A. Cammarata and G. K. Menon, *J. Med. Chem.*, **19**, 739 (1976).

(2) T. L. Isenhour and P. C. Jurs, *Anal. Chem.*, **43**, 20A (1971).

(3) B. R. Kowalski and C. F. Bender, *J. Am. Chem. Soc.*, **94**, 5632 (1972).

(4) B. R. Kowalski, *Anal. Chem.*, **47**, 1152A (1975).

(5) D. R. Preuss and P. C. Jurs, *ibid.*, **46**, 520 (1974).

(6) J. B. Justice and T. L. Isenhour, *ibid.*, **46**, 223 (1974).

(7) K. C. Chu, *ibid.*, **46**, 1181 (1974).

(8) K. C. Chu, R. J. Feldmann, M. B. Shapiro, G. F. Hazard, Jr., and R. I. Geran, *J. Med. Chem.*, **18**, 539 (1975).

(9) A. J. Stuper and P. C. Jurs, *J. Am. Chem. Soc.*, **97**, 182 (1975).

(10) B. R. Kowalski and C. F. Bender, *ibid.*, **96**, 916 (1974).

(11) H. C. Andrews, "Introduction to Mathematical Techniques in Pattern Recognition," Wiley–Interscience, New York, N.Y., 1972.

(12) W. S. Meisel, "Computer Oriented Approaches to Pattern Recognition," Wiley–Interscience, New York, N.Y., 1972.

(13) P. C. Jurs and T. L. Isenhour, "Chemical Applications of Pattern Recognition," Wiley–Interscience, New York, N.Y., 1975.

(14) C. Chen, "Statistical Pattern Recognition," Hayden Book Co., Rochelle Park, N.J., 1973.

(15) R. J. Mathews, *J. Am. Chem. Soc.*, **97**, 935 (1975).

(16) C. L. Perrin, *Science*, **183**, 551 (1974).

(17) K. H. Ting, R. C. T. Lee, G. W. A. Milne, M. Shapiro, and A. M. Guarino, *ibid.*, **180**, 417 (1973).

(18) R. D. Cramer, III, G. Redl, and C. E. Berkoff, *J. Med. Chem.*, **17**, 533 (1974).

(19) M. L. Weiner and P. H. Weiner, *ibid.*, **16**, 655 (1973).

(20) H. L. Friedman, *Symp. Chem. Biol. Correl., 1st Nat. Res. Council Publ.*, **206**, 296 (1951).

(21) L. L. Thurstone, "Multiple Factor Analysis," University of Chi-

cago Press, Chicago, Ill., 1947.
(22) S. A. Mulaik, "The Foundations of Factor Analysis," McGraw-Hill, New York, N.Y., 1972.
(23) J. E. Overall and C. J. Klett, "Applied Multivariate Analysis," McGraw-Hill, New York, N.Y., 1972.
(24) L. L. Thurstone, "The Vectors of the Mind," University of Chicago Press, Chicago, Ill., 1935.
(25) N. R. Bohidar, F. A. Restaino, and J. B. Schwartz, *J. Pharm. Sci.*, **64**, 966 (1975).
(26) C. Hansch, S. H. Unger, and A. B. Forsythe, *J. Med. Chem.*, **16**, 1217 (1973).
(27) S. R. Heller, C. L. Chang, and K. C. Chu, *Anal. Chem.*, **46**, 951 (1974).
(28) F. K. Kawahara, J. F. Santner, and E. C. Julian, *ibid.*, **46**, 266 (1974).
(29) N. J. Nilsson, "Learning Machines," McGraw-Hill, New York, N.Y., 1965.
(30) P. C. Jurs, B. R. Kowalski, T. L. Isenhour, and C. N. Reilley, *Anal. Chem.*, **41**, 690 (1969).
(31) B. R. Kowalski, P. C. Jurs, T. L. Isenhour, and C. N. Reilley, *ibid.*, **41**, 695 (1969).
(32) S. S. Schiffman, *Science*, **185**, 112 (1974).
(33) A. E. Taylor, "Advanced Calculus," Blaisdell, New York, N.Y., 1965.
(34) T. Fujita, J. Iwasa, and C. Hansch, *J. Med. Chem.*, **8**, 150 (1965).
(35) A. Cammarata, *ibid.*, **10**, 525 (1967).
(36) "Medicinal Chemistry," vol. II, 3rd ed., A. Burger, Ed., Wiley–Interscience, New York, N.Y., 1970.
(37) "AMA Drug Evaluations," 2nd ed., Publishing Sciences Group, Acton, Mass., 1973.
(38) A. I. Vogel, "Elementary Practical Organic Chemistry Part 2: Qualitative Organic Analysis," 2nd ed., Wiley, New York, N.Y., 1966.

ACKNOWLEDGMENTS AND ADDRESSES

Received December 23, 1975, from the *School of Pharmacy, Temple University, Philadelphia, PA 19140.*

Accepted for publication April 23, 1976.

The support provided G. K. Menon through a Temple University Graduate Fellowship is gratefully acknowledged.

[x] To whom inquiries should be directed.

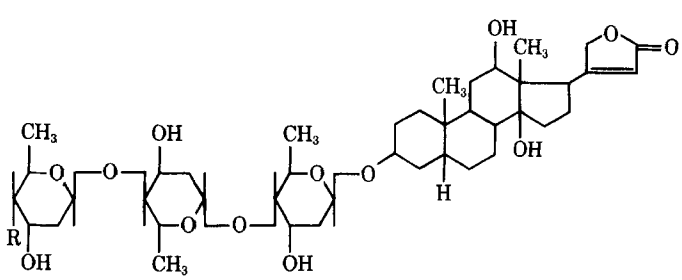
Pharmacokinetics of β-Methyldigoxin in Healthy Humans II: Oral Studies and Bioavailability

PETER H. HINDERLING [*], EDWARD R. GARRETT [x], and RONALD C. WESTER [‡]

Abstract □ The pharmacokinetics of orally administered aqueous ^{3}H-β-methyldigoxin solutions were studied at two dose levels, 0.3 and 0.6 mg, in healthy human subjects. The drug and its metabolites were specifically assayed in biological fluids and compared with results after intravenous doses to the same subjects. No significant dose dependency was observed. The apparent half-life of absorption was 16 ± 6 min (*SEM*). Digoxin was the only metabolite observed in the plasma and comprised 28.6 ± 3.7% of the dose in the urine. ^{3}H-β-Methyldigoxin, renally excreted unchanged, comprised 25.7 ± 1.7% (*SEM*). Water-soluble metabolites in the urine comprised 9.0 ± 1.8%. Fecal and urinary excretion accounted for 85% of the dose at 144 hr. The oral absorption of unchanged ^{3}H-β-methyldigoxin from solution was 59 ± 6% by area under the curve methods and 60 ± 4% by renal excretion. A total of 73% of the dose in the solution was absorbed as β-methyldigoxin and digoxin. First-pass metabolism prior to absorption was largely prehepatic and assignable to GI degradation; 21.9 ± 2.8% was degraded with 12.8 ± 4.0% to digoxin and 9.1 ± 4.0% to water-soluble metabolites. From 14 to 18% of the administered oral dose did not reach the systemic circulation. Analog computer fitting of plasma and urine levels of drug and digoxin was consistent with the first-pass premise with a delayed absorption of GI-generated digoxin and other metabolites. There were no significant differences between the oral absorption of a tablet formulation and the solution. Orally administered β-methyldigoxin solution delivered 97% cardioactivity as itself and digoxin with respect to an equivalent amount of intravenously administered digoxin. This value contrasts to the 140% delivered by intravenously administered β-methyldigoxin on the premise of pharmacodynamic equivalence of systemically appearing digoxin and β-methyldigoxin. Literature reports on the oral bioavailability of solutions and solid dosage forms of digoxin were critically reviewed, but no reliable comparison of the extent and reproducibility of oral absorption of cardioactive agents from administered digoxin or β-methyldigoxin could be made from the widely variable digoxin studies with nonspecific assays.

Keyphrases □ β-Methyldigoxin—oral, pharmacokinetics and bioavailability in humans, radiochemical–TLC study □ Pharmacokinetics—oral β-methyldigoxin, humans, radiochemical–TLC study □ Bioavailability—oral β-methyldigoxin, humans, radiochemical–TLC study □ Radiochemistry–TLC—study of pharmacokinetics and bioavailability of oral β-methyldigoxin in humans □ Cardiac glycosides—β-methyldigoxin, oral, pharmacokinetics and bioavailability in humans, radiochemical–TLC study

Semisynthetic derivatives of digoxin (Ia) such as β-methyldigoxin (Ib), a methyl ether of digoxin, and β-acetyldigoxin (Ic), an acetyl ester of digoxin, have been claimed to have higher intrinsic rate constants and efficiencies of absorption than digoxin in animals and in humans (1–14). Presumably, the rationale for their preferred usage is that a completely absorbed compound has the most consistent bioavailability in multiple dosage regimens. This is a valid approach, since glycosides have a narrow therapeutic range (15–18) and the occurrence of toxic manifestations in patients undergoing chronic therapy is 7–20% (19). Whereas β-acetyldigoxin was extensively metabolized or degraded before reaching the systemic circulation (20), β-methyldigoxin was claimed to



Ia: R = OH
Ib: R = OCH_3 Ic: R = $OCCH_3$