Quantitative Structure–Activity Relationships Employing Independent Quantum Chemical Indices

István Lukovits

Central Research Institute of Chemistry, Hungarian Academy of Sciences, H-1025 Budapest, Pusztaszeri ut 59, Hungary.
Received May 14, 1982

Derivation of quantitative structure–activity relationships between pharmacological potencies and the electronic structure of molecules may often result in chance correlations, because of the large number of quantum chemical indices. Interrelationships between the parameters complicate the interpretation of the results. Quantum chemical indices of benzylamines, tetracyclines, and 1,4-benzodiazepines were transformed into mutually independent components using principal component analysis. The number of essential components was 3, 4, and 3, respectively. The computational efforts needed to develop multivariate linear regression equations between these components and the pharmacological activities were reduced, since the regression coefficients were not affected by the inclusion of new parameters. In each example, the first component, which accounted for the highest part of the total sample variance in the electronic structure, was the most important one in determining pharmacological activity. It seems that besides the electrostatic forces, charge transfer also affected the inhibitory potencies of benzylamines.

Quantum chemical indices are used occasionally to predict the expected pharmacological activities of new therapeutic agents.[1] Because the computation of these indices may be time consuming and expensive and also because the results are more difficult to interpret and often less relevant than the results obtained using empirical substituent constants,[2] the latter are usually preferred to derive quantitative structure–activity relationships[3] (QSAR). It is assumed that the substituent constants will model adequately the changes in the physicochemical properties caused by substitution,[4] even if the chemical structure of the molecules is rather different from that of the reference compounds used to define these indices. This assumption is not necessary if direct experimental measurements are used to obtain the physicochemical indices or if quantum chemical calculations are done. Thus, quantum chemical indices may be useful in QSAR analysis, provided that both ground as well as excited state properties[5] are considered. The various indices may be closely related to each other. Because of these intercorrelations, one must be cautions in interpreting the meaning of the multiple regression equations calculated between the pharmacological activities and the quantum chemical indices of molecules.[6] A further problem arises if the number of pharmacological data, n, is small as compared to the number of indices, m. Increasing the number of parameters, the probability of finding change correlations will also increase.[7]

The aim of this paper is to demonstrate that the number of quantum chemical indices can be reduced, without losing information. Published results relating quantum chemical indices to pharmacological potencies were collected, and the quantum chemical indices were decomposed into mutually independent components. The term "independent" will mean in this paper that the correlation coefficient between the different parameters is zero. The transformation coefficients were calculated by using principal component analysis.[8]

Principal component analysis was used several times in QSAR studies. The investigations involved geometric descriptor indices, substituent constants,[9-13] pharmacological activities,[14,15] and the partition coefficient with the energy of the lowest molecular orbital.[16] In this paper, only quantum chemical indices were considered. A good introduction to principal components regression is given by Kendall.[17] Mason and Gunst investigated the problem of multicollinearity.[18,19] Malinowski and Howery,[20] Massy,[21] and Hill, Fomby and Johnson[22] investigated methods to determine the number of essential components.

The results of this study indicated that the number of essential factors is less than the number of primary indices and that the first component was the most important one in determining pharmacological activity.

Methods

The method of calculation was essentially the same, used in a previous paper.[14] The set of m primary quantum chemical indices of molecule i (i = 1, 2, ..., n), x_{i1}, x_{i2}, ..., x_{im}, form the $n \times m$ matrix $\mathbf{x}$. The actual values of variable j form the column vector $\mathbf{x}_j$ of this matrix (j = 1, 2, ..., m). Since the variables are not independent, the correlation coefficients calculated between them are not zero. The primary indices were adjusted to zero mean and unit standard deviation (eq 1), where $\bar{x}_j$ denotes the

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{1}$$

mean value, and s_j is the standard deviation of variable j. The transformed variables are dimensionless. This transformation *does not* alter the correlation coefficients. Following this procedure, a second transformation gives eq 2, where $\tilde{\mathbf{x}}$ denotes the

$$\mathbf{X} = \tilde{\mathbf{x}}\mathbf{c} \tag{2}$$

matrix of normalized primary indices, c denotes the $n \times m$ matrix of transformation coefficients, and the $n \times m$ matrix $\mathbf{X}$ consists

(1) Schnaare, R. S. *Drug Des.* **1971**, *1*, 405.
(2) Hansch, C.; Leo, A.; Unger, S. H.; Kim, K. H.; Nikaitani, D.; Lien, E. J. *J. Med. Chem.* **1973**, *16*, 1207.
(3) Martin, Y. C. *J. Med. Chem.* **1981**, *24*, 229.
(4) Hansch, C. *J. Chem. Educ.* **1974**, *51*, 360.
(5) Klopman, G.; Hudson, R. F. *Theor. Chim. Acta* **1967**, *8*, 165.
(6) Martin, Y. C. *J. Med. Chem.* **1970**, *13*, 145.
(7) Topliss, J. G.; Costello, R. J. *J. Med. Chem.* **1972**, *15*, 1066.
(8) Anderson, T. W. "Introduction to Multivariate Statistical Analysis"; Wiley: New York, 1970.
(9) Franke, R. "Optimierungsmethoden in der Wirkstofforschung–Quantitative Struktur–Wirkungs–Analyse"; Akademie-Verlag: Berlin, 1980.
(10) Cammarata, A.; Menon, G. K. *J. Med. Chem.* **1976**, *19*, 739.
(11) Dunn III, W. J.; Wold, S.; Martin Y. C. *J. Med. Chem.* **1978**, *21*, 922.
(12) Dunn III, W. J.; Wold, S. *J. Med. Chem.* **1978**, *21*, 1001.
(13) Streich, W. J.; Dove, S.; Franke, R. *J. Med. Chem.* **1980**, *23*, 1452.
(14) Lukovits, I.; Lopata, A. *J. Med. Chem.* **1980**, *23*, 449.
(15) Weiner, M. L.; Weiner, P. H. *J. Med. Chem.* **1973**, *16*, 655.
(16) Dunn III, W. J.; Wold, S. *J. Med. Chem.* **1980**, *23*, 595.
(17) Kendall, M. G. "A Course in Multivariate Analysis"; Charles Griffin and Co.: London, 1957.
(18) Gunst, R. F.; Mason, R. L. *Biometrics* **1977**, *33*, 249.
(19) Mason, R. L.; Gunst, R. F.; Webster, J. T. *Commun. Stat.* **1975**, *4*, 277.
(20) Malinowski E. R.; Howery, D. G. "Factor Analysis in Chemistry"; Wiley: New York, 1980.
(21) Massy, W. F. *J. Am. Stat. Assoc.* **1965**, *60*, 234.
(22) Hill, R. C.; Fomby, T. B.; Johnson, S. R. *Commun. Stat.-Theor. Methods* **1977**, *A6*, 309.

0022-2623/83/1826-1104$01.50/0

Table I. Principal Components of Quantum Chemical Indices of Benzylamines (I) and Pharmacological Potencies[a]

no.[b]	substituents	X_1^I (45.3)	X_2^I (22.8)	X_3^I (15.5)	activities[c] obsd	activities[c] estd (eq 16)
1	2-Cl, 3-CF_3	−3.29	0.25	−0.22	6.82	5.74
2	2,3-Cl_2	−3.04	0.83	−0.64	6.23	5.52
3	3-CF_3	−1.04	−3.00	0.22	5.64	5.45
4	4-CF_3	−0.08	−2.11	1.88	5.20	4.87
5	3-Cl	−0.78	−0.09	−0.20	5.07	4.76
6	3,4-Cl_2	−3.09	0.42	1.15	4.97	5.62
7	2,5-Cl_2	−1.23	0.97	−0.55	4.82	4.73
8	2,4-Cl_2	−2.17	0.25	0.40	4.68	5.27
9	2-Cl	−1.01	1.15	−1.35	4.66	4.60
10	2-CF_3	0.40	−2.00	−1.10	4.64	4.64
11	2,3-$(CH_3)_2$	1.99	0.73	−0.18	4.24	3.43
12	2-CH_3	1.67	1.17	−0.41	4.17	3.47
13	2,6-Cl_2	−0.15	0.54	−2.03	4.08	4.37
14	4-Cl	−0.31	−0.02	1.02	4.07	4.54
15	3-F	1.63	−2.14	0.79	4.00	4.15
16	3,5-Cl_2	−1.65	0.15	0.52	3.93	5.07
17	2-F	1.84	0.16	−1.14	3.87	3.61
18	3-CH_3	2.46	−0.73	−0.38	3.82	3.52
19	4-F	1.59	0.22	1.06	3.17	3.70
20	4-CH_3	2.01	0.92	−0.94	3.16	3.38
21	H	2.20	−0.64	−0.58	3.12	3.61
22	4-OCH_3	2.05	2.99	2.67	2.78	2.95

[a] The numbers in parentheses denote the percentage of the total sample variance explained by the variables. [b] The quantum chemical indices were taken from ref 25. [c] pI_{50} (in moles per liter).

Table II. Principal Components of Quantum Chemical Indices of Tetracyclines (II) and Pharmacological Potencies[a]

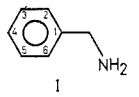
no.[b]	tetracycline	X_1^{II} (36.2)	X_2^{II} (25.5)	X_3^{II} (18.3)	X_4^{II} (11.8)	exptl act.[c]
1	7-nitro-6-demethyl-6-deoxy-	2.35	3.82	−1.08	0.06	2.87
2	7-chloro-6-demethyl-	0.52	−0.68	0.11	0.09	2.71
3	7-chloro-	0.55	−0.83	−0.02	0.36	2.60
4		0.61	−0.70	−0.06	0.40	2.43
5	5-oxy-	0.60	−0.66	−0.03	0.33	2.41
6	7-amino-6-demethyl-6-deoxy-	0.20	−0.67	0.73	−1.05	2.26
7	9-amino-6-demethyl-6-deoxy-	0.14	−0.73	0.77	−1.07	2.16
8	6-demethyl-6-deoxy-	0.48	−0.07	0.48	−0.76	1.98
9	7-bromo-6-demethyl-6-deoxy-	0.35	−0.16	0.70	−0.39	1.71
10	9-nitro-6-demethyl-6-deoxy-	1.19	2.58	−0.04	−0.09	1.65
11	9-(dimethylamino)-6-demethyl-6-deoxy-	−0.02	−1.43	0.97	−1.26	1.37
12	5a(6)-anhydro-	−6.12	1.89	0.40	0.02	1.19
13	12a-deoxy-	0.12	−0.73	1.31	3.21	0.41
14	7-chloro-5a(11a)-deoxy-	−0.96	−1.62	−4.25	0.16	−0.47

[a] See footnote a in Table I. [b] The quantum chemical indices were taken from ref 26. [c] Logarithms of the inhibitory rate constants (mol^{-1} s^{-1}) against *Escherichia coli*.

of column vectors X_j, which are independent. The coefficients c_{ij} forming matrix c, were obtained by using principal component analysis.[8] The sum of the sample variances of the transformed indices $\tilde{x}_j$ is m and is referred to as *total sample variance*. The sum of the variances of variables X_j is also m. While the variance of each variable $\tilde{x}_j$ is 1, the variances of variables X_j are different. In this paper, X_1, the first component, denotes the variable with the highest variance, X_2, the second component, denotes the variable with the second highest variance, etc. Since these variances are the eigenvalues of the matrix of intercorrelation coefficients, the two terms will be used interchangeably.

The determination of the essential components caused some problems, since most methods used in factor analysis require the knowledge of the experimental error of the indices.[20] Components having low variance are representing the experimental error and the random noise. These factors should be deleted. No "experimental error" can be attributed to quantum chemical indices. The inbedded error function, the factor indicator function,[20,23] and the "eigenvalue larger than one" criterion[24] could be tried here, since these methods do not require the information on the experimental error. The "eigenvalue larger than one" criterion was used in this paper, meaning that all eigenvalues and corresponding components above the average eigenvalue were accepted and those below this value were rejected. If the matrix of correlation coefficients is diagonalized, the threshold value is 1.

The quantum chemical indices considered here were taken from the papers of Otto et al.[25] (set I), Peradejordi, Martin, and Cammarata[26] (set II), and Blair and Webb[27] (set III). Set I consists



I

of 22 benzylamine (I) derivatives (Table I). The inhibitory activities against phenylethanolamine *N*-methyltransferase (PNMT) were tested by Fuller et al.[28] The CNDO/2 method was used to calculate the charge density (q) at atom 1 (q_1), the sum of atomic charges at atoms 2–6 of the benzene ring ($\sum q'$, where the prime indicates that atom 1 is not added), the dipole moment (μ) and its orthogonal components (μ_x and μ_y), the first

(23) Malinowski, E. R. *Anal. Chem.* **1977**, *49*, 612.
(24) Kaiser, H. F. *Educ. Psychol. Meas.* **1960**, *20*, 141.
(25) Otto, P.; Seel, M.; Ladik, J.; Müller, P. *J. Theor. Biol.* **1979**, *78*, 197.
(26) Peradejordi, F.; Martin, A. N.; Cammarata, A. *J. Pharm. Sci.* **1971**, *60*, 576.
(27) Blair, T.; Webb, G. A. *J. Med. Chem.* **1977**, *20*, 1206.
(28) Fuller, R. M.; Molloy, B. B.; Day, W. A.; Rouch, B. W.; March, M. M. *J. Med. Chem.* **1973**, *16*, 101.

Table III. Principal Components of Quantum Chemical Indices of 1,4-Benzodiazepin-2-one Derivatives and Pharmacological Potencies[a]

no.[b]	R_1	R_2	R_2'	X_1^{III} (49.4)	X_2^{III} (39.3)	X_3^{III} (8.7)	activities[c] obsd	activities[c] estd (eq 25)
1	H	H	H	3.04	−0.76	−0.03	0.37	0.41
2	H	F	H	1.87	0.60	−0.84	0.40	0.70
3	H	Cl	H	1.84	1.15	0.11	1.13	1.14
4	H	CN	H	1.93	0.84	0.41	0.81	1.16
5	H	NO_2	H	0.66	2.92	−0.03	1.75	1.75
6	H	CF_3	H	1.38	1.72	0.03	1.48	1.34
9	H	SCH_3	H	2.68	−0.19	−0.07	1.15	0.60
13	CH_3	Cl	H	−1.44	−0.50	−0.12	1.45	1.56
14	CH_3	NO_2	H	−2.09	0.89	−0.08	2.07	2.01
15	CH_3	CN	H	−1.38	−0.54	0.20	1.74	1.65
16	CH_3	SCH_3	H	−0.42	−1.87	−0.20	0.87	1.00
17	CH_3	$N(CH_3)_2$	H	−0.15	−2.63	−0.80	0.86	0.57
18	CH_3	Cl	F	−1.12	−0.69	1.13	1.78	1.86
19	CH_3	Cl	Cl	−1.88	0.60	−0.68	1.50	1.70
20	CH_3	NO_2	F	−1.88	0.69	1.20	2.59	2.36
21	CH_3	NO_2	Cl	−2.41	1.78	−0.76	2.12	2.04
22	CH_3	NO_2	CF_3	−0.65	0.58	0.80	1.86	1.89
23	CH_3	H	F	0.60	−2.94	1.15	0.83	0.99
25	CH_3	$N(CH_3)_2$	Cl	−0.57	−1.64	−1.42	0.61	0.66

[a] See footnote *a* in Table I. [b] The quantum chemical indices were taken from ref 27. [c] Log ED_{50} (in millimoles per kilogram) obtained by using the foot-shock test in mice.[29]

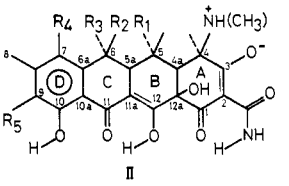
Table IV. Transformation Equations[a]

equation	no.
$X_1^I = 0.25q_1 + 0.37\Sigma q' - 0.42\mu + 0.24\mu_x + 0.36\mu_y + 0.44\Delta E - 0.02\epsilon_{HOMO} + 0.49\epsilon_{LEMO}$	3
$X_2^I = 0.55q_1 - 0.27\Sigma q' - 0.02\mu - 0.04\mu_x + 0.23\mu_y - 0.33\Delta E - 0.67\epsilon_{HOMO} + 0.08\epsilon_{LEMO}$	4
$X_3^I = 0.07q_1 + 0.30\Sigma q' + 0.37\mu - 0.78\mu_x + 0.38\mu_y + 0.09\Delta E - 0.07\epsilon_{HOMO} + 0.08\epsilon_{LEMO}$	5
$X_1^{II} = 0.07q_6 + 0.14q_{10} + 0.27q_{11} - 0.48q_{12} - 0.26E_{10} + 0.26N_{10} - 0.15E_{11} - 0.45N_{11} + 0.40E_{12} + 0.39N_{12}$	6
$X_2^{II} = -0.32q_6 + 0.55q_{10} - 0.32q_{11} + 0.20q_{12} - 0.46E_{10} + 0.43N_{10} + 0.02E_{11} + 0.22N_{11} - 0.002E_{12} - 0.01N_{12}$	7
$X_3^{II} = -0.27q_6 - 0.17q_{10} - 0.45q_{11} - 0.06q_{12} + 0.21E_{10} - 0.16N_{10} + 0.33E_{11} + 0.22N_{11} + 0.47E_{12} + 0.49N_{12}$	8
$X_4^{II} = 0.59q_6 + 0.15q_{10} + 0.10q_{11} - 0.13q_{12} - 0.23E_{10} - 0.004N_{10} + 0.71E_{11} + 0.21N_{11} - 0.05E_{12} + 0.03N_{12}$	9
$X_1^{III} = -0.51q_1 + 0.42q_2 + 0.55q_3 - 0.23q_4 - 0.28q_0 + 0.36\mu$	10
$X_2^{III} = -0.28q_1 + 0.44q_2 + 0.15q_3 + 0.44q_4 + 0.54q_0 - 0.46\mu$	11
$X_3^{III} = 0.08q_1 + 0.07q_2 + 0.08q_3 - 0.85q_4 + 0.33q_0 - 0.39\mu$	12

[a] See footnote *b* in Tables I–III. These equations are related to normalized quantum chemical indices.

singlet excitation energy (ΔE), and the energy of the highest occupied (ϵ_{HOMO}) as well as of the lowest empty molecular orbital (ϵ_{LEMO}).

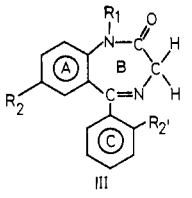
Set II consists of 14 tetracycline (II) derivatives (Table II).



II

Peradejordi et al.[26] determined the inhibitory rate constants (k) against *Escherichia coli*. Using the simple Hückel approach, the authors calculated the charge densities of atoms 6, 10, 11, and 12 as well as the electrophilic (E) and nucleophilic (N) delocalizabilities of atoms 10–12.

Set III consists of 19 1,4-benzodiazepin-2-one derivatives (III),



III

the pharmacological activities of which (Table III) were determined by using the foot-shock test.[29] Blair and Webb[27] used the CNDO/2 method to calculate the charge densities of atoms 1–4, the carbonyl oxygen (q_O), and the dipole moment.

The pharmacological potencies are denoted by A^Y, where the upper index Y is used to specify the set {Y = I, II, III}. The same index is used to specify the principal components X_j^Y (Tables IV and V).

Results

The principal components of the total sample variance of set I were 3.62, 1.82, 1.24, 0.66, 0.36, 0.17, 0.07, and 0.04. According to the "eigenvalue larger than one" criterion, only the first three components were essential. These components are listed in Table I. The three components accounted for 83% of the total sample variance. Equations 3–5 (Table IV) define these components in terms of the primary normalized indices. The regression of the activities A^I on these components are expressed by eq 13–15 (Table V). Here R^2 denotes the square of the correlation coefficient, s is the standard error of the estimate, and F denotes the result of the F test.[30] Numbers in parentheses

(29) Sternbach, L. H.; Randall, L. O.; Banziger, R.; Lehr, H. In "Drugs Affecting the Central Nervous System", Berger, A., Ed.; Marcel Dekker: New York, 1973; Vol. 2.

Table V. Regression Equations Calculated between Pharmacological Potencies and the Principal Components of Quantum Chemical Indices[a]

equation	n	R^2	s	F	no.
Benzylamines (I)					
$A^{I} = -0.42\ (\pm 0.15)\ X_1^{I} + 4.42$	22	0.632	0.62	34.38	13
$A^{I} = -0.20\ (\pm 0.33)\ X_2^{I} + 4.42$	22	0.072	0.99	1.56	14
$A^{I} = -0.12\ (\pm 0.42)\ X_3^{I} + 4.42$	22	0.018	1.02	0.37	15
$A^{I} = -0.42\ (\pm 0.14)\ X_1^{I} - 0.20\ (\pm 0.19)\ X_2^{I} + 4.42$	22	0.705	0.57	22.66	16
Tetracyclines (II)					
$A^{II} = 0.21\ (\pm 0.28)\ X_1^{II} + 1.81$	14	0.184	0.88	2.70	17
$A^{II} = 0.16\ (\pm 0.36)\ X_2^{II} + 1.81$	14	0.070	0.94	0.90	18
$A^{II} = 0.28\ (\pm 0.40)\ X_3^{II} + 1.81$	14	0.164	0.89	2.36	19
$A^{II} = -0.29\ (\pm 0.51)\ X_4^{II} + 1.81$	14	0.109	0.92	1.47	20
$A^{II} = 0.21\ (\pm 0.26)\ X_1^{II} + 0.28\ (\pm 0.36)\ X_2^{II} - 0.29\ (\pm 0.45)\ X_4^{II} + 1.81$	14	0.457	0.79	2.11	21
Benzodiazepines (III)					
$A^{III} = -0.25\ (\pm 0.13)\ X_1^{III} + 1.34$	19	0.483	0.46	15.87	22
$A^{III} = 0.20\ (\pm 0.18)\ X_2^{III} + 1.34$	19	0.246	0.56	3.91	23
$A^{III} = 0.34\ (\pm 0.40)\ X_3^{III} + 1.34$	19	0.158	0.59	3.20	24
$A^{III} = -0.25\ (\pm 0.07)\ X_1^{III} + 0.20\ (\pm 0.08)\ X_2^{III} + 0.34\ (\pm 0.16)\ X_3^{III} + 1.34$	19	0.887	0.23	39.26	25

[a] See footnote *c* in Tables I-III.

denote the 95% confidence intervals of the regression coefficients. Equation 13 is significant at the $p < 0.01$ level, since the calculated F value was much higher than the theoretical one ($F_{1,20,p=0.01} = 8.10$). Components X_2^{I} and X_3^{I} were inferior than X_1^{I} (eq 14 and 15), but consideration of X_2^{I} together with X_1^{I} improved the multiple correlation coefficient significantly at the $p < 0.05$ level, since the partial F test[30] yielded a value $F_{1,19} = 4.46$ against the theoretical value $F_{1,19,p=0.5} = 4.38$ (eq 16). Addition of X_3^{I} did not improve this correlation ($F_{1,18} = 1.17$ vs. $F_{1,18,p=0.25} = 1.41$). The overall F statistics (Table V) indicated that eq 16 was significant at the $p < 0.01$ level ($F_{3,18,p=0.01} = 5.09$). The multiple correlation coefficient is the Pythagorean sum of the simple correlation coefficients,[8] $R = 0.704^{1/2} = 0.839 = (0.632 + 0.072)^{1/2}$ (eq 13, 14, and 16).

The principal components of the total sample variance of set II were 3.62, 2.55, 1.83, 1.18, 0.50, 0.27, 0.05, 0.004, 0.002, and 8×10^{-8}. The first four components that were essential according to the "eigenvalue larger than one" criterion are listed in Table II. These components explained 92% of the total sample variance. Equations 6–9 (Table IV) are defining these components in terms of the normalized primary indices. The regressions of the experimental potency (A^{II}) on these components are given by eq 17–20 (Table V). The correlation coefficients were rather low; neither of them was significant at the $p < 0.05$ level ($F_{1,12,p=0.05} = 4.75$). Combination of X_1^{II}, X_2^{II}, and X_4^{II} explained less than 50% of the variation in the dependent variable A^{II} (eq 21), and the equation was not significant at the $p < 0.05$ level ($F_{3,10,p=0.05} = 3.71$). Equation 21 could not be used for quantitative predictions.

The principal components of the total sample variance of set III were 2.97, 2.36, 0.52, 0.08, 0.06, and 0.02. According to the "eigenvalue larger than one" criterion, only the first two components were essential, but it could be shown that the third component was still important in determining pharmacological activity (see below). The first three components X_1^{III}, X_2^{III}, and X_3^{III} are defined by eq 10–12 (Table IV), and the regressions of the experimental activities on these components are given by eq 22–24 (Table V). No multiple linear regression equation including all six quantum chemical indices was developed, but it could be shown that the multiple correlation coefficient would be in this case $R = 0.948$. The first components (eq 25) practically reproduced this figure ($R = 0.942$); the difference was not significant even at the $p < 0.1$ level ($F_{1,14} = 1.67$ vs. $F_{1,14,p=0.1} = 3.10$). The first component was the "best" one; 48% of the variation in the pharmacological potencies was associated with X_1^{III} (eq 22). Consideration of X_2^{III} improved eq 22 significantly (eq 23) at the $p < 0.01$ level, as indicated by the partial F test ($F_{1,16} = 14.49$ vs. $E_{1,16,p=0.01} = 8.53$); inclusion of X_3^{III} (eq 24) again improved the correlation coefficient significantly ($F_{1,15} = 21.04$ vs. $F_{1,15,p=0.01} = 8.68$).

The results are summarized by eq 26–28. The coefficients are the sums of products of the transformation coefficients (Table IV) and the regression coefficients in eq 16, 21, and 25.

$$A^{I} = -0.22q_1 - 0.10\sum q' + 0.18\mu - 0.09\mu_x - 0.20\mu_y - 0.12\Delta E + 0.15\epsilon_{HOMO} - 0.22\epsilon_{LEMO} \quad (26)$$

$$A^{II} = -0.28q_6 + 0.03q_{10} - 0.15q_{11} - 0.05q_{12} + 0.08N_{10} - 0.14E_{11} - 0.06N_{11} + 0.23E_{12} + 0.21N_{12} \quad (27)$$

$$A^{III} = 0.10q_1 + 0.01q_2 - 0.08q_3 - 0.14q_4 + 0.29q_6 - 0.31\mu \quad (28)$$

Discussion

Application of principal components regression to quantum chemical indices reduced the computational efforts, since the order of parameters to be included into the multiple linear regression equations was determined by the absolute magnitude of the correlation coefficients calculated between the dependent variable and the respective components. Often, to get better results, all combinations in the parameters are tried in QSAR studies. Because of the large number of quantum chemical indices, thousands[31] or ten thousands[32] of combinations of these indices had to be investigated. This "walking around the data"[33] could be avoided by using principal components regression. The

(30) Snedecor, G. W.; Cochran, W. G. "Statistical Methods"; The Iowa State University Press: Ames, 1972.
(31) Sklenar, H.; Jäger, J. *Int. J. Quantum Chem.* **1979**, *16*, 467.
(32) Scharfenberg, P.; Sauer, J. *Int. J. Quantum Chem.* **1980**, *18*, 1309.
(33) Hansch, C.; Silipo, C. *J. Med. Chem.* **1974**, *17*, 661.

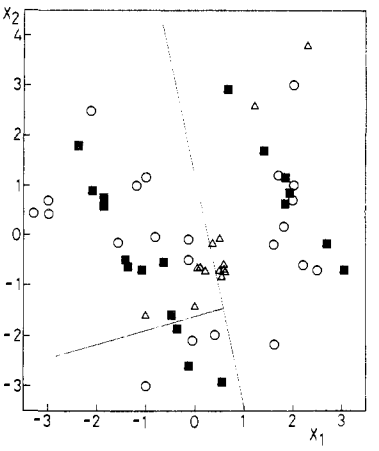


Figure 1. Plot of the first against the second principal component of the quantum chemical indices of benzylamines (○), tetracycline derivatives (△), and benzodiazepines (■). The lines separate the subgroups of benzylamines.

development of eq 16, 21, and 25 was easy, since the regression coefficients were not changed by the inclusion of new parameters.

In all examples, the first components were the most important ones in determining the pharmacological potencies.

Otto et al.[25] had to try $\binom{8}{1} + \binom{8}{2} = 36$ monovariate and bivariate combinations in the parameters and obtained $R = 0.88$ for μ_x and μ_y. Based on this result, the authors suggested that electrostatic interactions may affect the pharmacological potencies of benzylamines. According to eq 26, all parameters affected to some extent the pharmacological response; q_1 ϵ_{LEMO}, and μ_y seemed to be the most important ones. The presence of index ϵ_{LEMO} in eq 26 indicated that charge transfer may also influence activity. Although the correlation coefficient (r) between A^I and μ_y was −0.83, whereas the correlation between A^I and ϵ_{LEMO} was −0.71, the former higher value may be a result of chance effects, since the coefficients of μ_y and of ϵ_{LEMO} were nearly equal in eq 26.

For the series of tetracyclines, Peradejordi et al.[26] considered 356 different combinations in the quantum chemical indices. The best regression equation involved seven parameters, and the multiple correlation coefficient was $R = 0.993$. The results presented in this paper did not confirm the conclusions made by Peradejordi et al.[26] The multiple correlation coefficient obtained by considering the essential components was much lower and was not significant. Thus, the value of $R = 0.993$ may be a result of chance correlation, in accordance with the suggestion made by Topliss and Costello.[7] Substitutions at saturated rings might cause local changes in the electronic structure, which could be simulated by using dummy variables. Although dummy variables may be used in QSAR analysis[34] and also in factor analysis,[20] the ones treated by Peradejordi et al.[26] might be closely related.

For the benzodiazepine derivatives, Blair and Webb[27] found that q_0 and μ are the most important indices in determining pharmacological activity. Our results indicated (eq 28) that variations in the charge densities (q_1, q_3, and q_4) might also affect the experimental activities.

The plots of the components X_2^Y against X_1^Y (Y = I–III) showed that the data points representing the benzylamines (Y = I) and the benzodiazepine (Y = III) derivatives were grouped (Figure 1). As for the benzylamines, there were three groups that corresponded to the chlorine derivatives, the methyl, fluoro, and trifluoromethyl derivatives. Within each group, X_1^I and X_2^I were independent. This is a balanced data set. Two groups could be identified for the benzodiazepines, the first one corresponds to 1-methyl derivatives and second one to derivatives containing H in position 1. Within both groups, X_2^{III} was linearly related to X_1^{III}. The centroids of the groups are relatively far away, what might contribute to the success of principal component analysis in this series. The tetracycline data set was not grouped, but after the data point related to molecule 12 in Table II was removed, the components X_1^{II} and X_2^{II} became linearly dependent (Figure 1). This means that the first component may be more important in explaining the dependent variable than indicated by eq 17.

Principal components regression should not be used if more or less independent sets of substituent constants are used to derive QSAR.[13,35-37] Principal components regression employing quantum chemical or experimental physicochemical indices may be used to derive QSAR if the molecular mechanism of the drug–receptor interaction is not known and none of the theoretical indices is an adequate parameter. The importance of the experimental indices has to be emphasized, since many of them, like the partition coefficient, cannot be calculated exactly.[38] Experimental indices could be combined with quantum chemical indices in principal components regression.

For sets I and II the number of *essential* components as determined by the "eigenvalue larger than one" criterion was higher than the number of components that were *important* in explaining pharmacological activity. However, it seems that the "eigenvalue larger than one" criterion should be used cautiously, since factors still important may be omitted (set III).

In summarizing this discussion it can be concluded that principal components regression can be used to derive QSAR if many strongly interrelated variables are considered, since it defines the number of significant, independent components, being less than the number of primary indices.

Acknowledgment. Thanks are due to Dr. G. Náray-Szabó for reading and commenting on the manuscript and to Dr. T. Šolmajer (Ljubljana) for helpful discussions. This work was supported by Grant KKP-3 3.2.3. from the Hungarian Academy of Sciences.

Registry No. I (2-Cl, 3-CF_3), 39226-96-5; I (2,3-Cl_2), 39226-95-4; I (3-CF_3), 2740-83-2; I (4-CF_3), 3300-51-4; I (3-Cl, 4152-90-3; I (3,4-Cl_2), 102-49-8; I (2,5-Cl_2), 10541-69-2; I (2,4-Cl_2), 95-00-1; I (2-Cl), 89-97-4; I (2-CF_3), 3048-01-9; I [2,3-$(CH_3)_2$], 51586-20-0; I (2-CH_3), 89-93-0; I (2,6-Cl_2), 6575-27-5; I (4-Cl), 104-86-9; I (3-F), 100-82-3; I (3,5-Cl_2), 39989-43-0; I (2-F), 89-99-6; I (3-CH_3), 100-81-2; I (4-F), 140-75-0; I (4-CH_3), 104-84-7; I, 100-46-9; I (4-OCH_3), 2393-23-9; II (R_1,R_2,R_3,R_5 = H; R_4 = NO_2), 5585-59-1; II (R_1,R_3,R_5 = H; R_2 = OH; R_4 = Cl), 127-33-3; II (R_1,R_5 = R_2 = OH; R_3 = CH_3; R_4 = Cl), 57-62-5; II (R_1,R_4,R_5 = H; R_2 = OH; R_3 = CH_3), 60-54-8; II (R_1,R_2 = OH; R_3 = CH_3; R_4,R_5 = H), 79-57-2; II (R_1,R_2,R_3,R_5 = H; R_4 = NH_2), 5679-00-5; II (R_{1-4} = H; R_5 = NH_2), 5874-95-3; II (R_{1-5} = H), 808-26-4; II (R_1,R_2,R_3,R_5 = H; R_4 = Br), 31642-30-5; II (R_{1-4} = H; R_5 = NO_2), 4199-35-3; II [R_{1-4} = H; R_5 = $N(CH_3)_2$], 10118-89-5; II [R_1,R_4,R_5 = H; R_3 = CH_3; 5a(6)-anhydro], 1665-56-1; II (R_1,R_4R_5 = H; R_2 = OH; R_3 = CH_3; 12a-deoxy), 4199-36-4; III (R_1,R_2,R_2' = H), 2898-08-0; III (R_1,R_2'

(34) Free, S. M.; Wilson, J. M. *J. Med. Chem.* **1964**, *7*, 395.
(35) Hansch, C.; Unger, S. H.; Forsythe, A. B. *J. Med. Chem.* **1973**, *16*, 1217.
(36) Wootton, R.; Cranfield, R.; Sheppey, G. C. *J. Med. Chem.* **1975**, *18*, 607.
(37) Wooldrige, K. R. H. *Eur. J. Med. Chem.-Chim. Ther.* **1980**, *15*, 63.
(38) Cammarata, A.; Rogers, K. S. *J. Med. Chem.* **1971**, *14*, 269.

= H; R_2 = F), 2648-00-2; III (R_1,$R_{2'}$ = H; R_2 = Cl), 1088-11-5; III (R_1,$R_{2'}$ = H; R^2 = CN), 17562-53-7; III (R_1,$R_{2'}$ = H; R_2 = NO_2), 146-22-5; III (R_1,$R_{2'}$ = H; R_2 = CF_3), 2285-16-7; III (R_1,$R_{2'}$ = H; R_2 = SCH_3), 2891-12-5; III (R_1 = CH_3; R_2 = Cl; $R_{2'}$ = H), 439-14-5; III (R_1 = CH_3; R_2 = NO_2; $R_{2'}$ = H), 2011-67-8; III (R_1 = CH_3; R_2 = CN; $R_{2'}$ = H), 3489-59-6; III (R_1 = CH_3; R_2 = $N(CH_3)_2$; $R_{2'}$ = H), 2891-09-0; III (R_1 = CH_3; R_2 = Cl; $R_{2'}$ = F), 3900-31-0; III (R_1 = CH_3; R_2,$R_{2'}$ = Cl), 2894-68-0; III (R_1 = CH_3; R_2 = NO_2; $R_{2'}$ = F), 1622-62-4; III (R_1 = CH_3; R_2 = NO_2; $R_{2'}$ = Cl), 5527-71-9; III (R_1 = CH_3; R_2 = NO_2; $R_{2'}$ = CF_3), 1959-37-1; III (R_1 = CH_3; R_2 = H; $R_{2'}$ = F), 844-11-1; III [R_1 = CH_3; R_2 = $N(CH_3)_2$; $R_{2'}$ =Cl], 30144-75-3; PNMT, 9037-68-7.

Ultra-Short-Acting β-Adrenergic Receptor Blocking Agents. 3. Ethylenediamine Derivatives of (Aryloxy)propanolamines Having Esters on the Aryl Function

Paul W. Erhardt,* Chi M. Woo,† William L. Matier,‡ Richard J. Gorczynski,§ and William G. Anderson⊥

American Critical Care, McGaw Park, Illinois 60085. *Received December 10, 1982*

Various ethylenediamine derivatives have been incorporated into the nitrogen substituent of certain short-acting (aryloxy)propanolamine systems that contain esters on their aryl functions. Although several of these compounds showed durations of action comparable to their prototypes, most of the nitrogen substituents significantly prolonged the duration of β-adrenergic blockade. Similarly, while one of the compounds showed appreciable cardioselectivity in vitro, generally, little enhancement of cardioselectivity was obtained. A brief discussion of structure–activity relationships observed for the ethylenediamine derivatives is presented.

As part of a program intended to produce short-acting β-adrenergic receptor blocking agents,[1] we previously described the syntheses and pharmacology of several (aryloxy)propanolamines that contained ester moieties incorporated into their nitrogen substituent[2] and aryl functions.[3] These studies revealed that when certain ester and alkyl ester groups are placed on the aryl function, short-acting β-blockers can be obtained. Furthermore, when the aryl substitution pattern is para, compounds exhibit moderate β-blocking potency and tend to be cardioselective, whereas potent, nonselective compounds result from ortho substitution (e.g., compounds 1 and 2 in Table I). An attempt to obtain both high potency and cardioselectivity by combining an *o*-carboalkoxy system with the 3,4-dimethoxyphenethyl nitrogen substituent[4] (compound 3) caused the duration of action to increase considerably.

Our interest in obtaining cardioselective short-acting compounds was two-fold: first, potential problems associated with β_2-blockade in the airways should be less; and, second, β_1-selectivity would decrease the possibility that the balance between α- and β_2-adrenergic control could be disrupted in favor of α-induced vasoconstriction in the coronary vasculature. The latter possibility could result in coronary vasospasm[5] and exacerbate certain of the cardiovascular pathologies[6] envisioned for therapy with an ultra-short-acting β-blocker.[1]

In this report we describe several additional potent *o*-carboalkoxy derivatives where a variety of ethylenediamine moieties[7-9] have been employed in an attempt to enhance cardioselectivity. These compounds are listed in Table I.

Chemisty. The syntheses of the test compounds followed established procedures. Methyl or ethyl salicylate was treated with epichlorohydrin,[2,3] providing key intermediate epoxides, which were then opened with the appropriate diamines to produce **4–20**. The diamines required for **16–20** and **22–24** were commercially available, and those for **4–15** were prepared by literature methods.[8-10] Compounds **22–24** were prepared starting from methyl 4-hydroxybenzoate, compound **27** was prepared starting from 2-methylphenol, and compound **25** was prepared starting from *o*-hydroxycinnamic acid after esterification with methanol.[2] Compound **26** was obtained from **25** by catalytic hydrogenation. The preparations of **1–3**, **21**, and **28** have been described previously.[2,3]

Chemical data for the test compounds are listed in Table I. A representative synthesis is provided under Experimental Section.

Results and Discussion

Compounds were first screened in vitro in guinea pig right atrial and tracheal preparations.[2] Compounds showing pA_2 values equal to or greater than 7.0 were then tested for their duration of effect in a canine preparation following a 3-h intravenous infusion. Infusion rates were adjusted to produce approximately 50% inhibition of isoproterenol-induced tachycardia. Isoproterenol challenges were performed at 10-min intervals before, during, and after the infusion of test compounds.

In the first series, amides **4–7**, a very small increase in cardioselectivity compared to **1** was obtained based on in vitro data. These compounds possessed durations of action somewhat longer than our desired value of approximately 10 min.[2,3] Cardioselectivity was not observed for sulfon-

* Corresponding address: Section of Medicinal Chemistry, Berlex Laboratories, Inc., Cedar Knolls, NJ 07927.
† Section of Medicinal Organic Chemistry, Department of Pharmaceutical Research.
‡ Department of Pharmaceutical Development.
§ Section of Pharmacology, Department of Pharmaceutical Research.
⊥ Current address: Department of Pharmacology, William H. Rorer, Fort Washington, PA 19034.

(1) Zaroslinski, J.; Borgman, R. J.; O'Donnell, J. P.; Anderson, W. G.; Erhardt, P. W.; Kam, S.-T; Reynolds, R. D.; Lee, R. J.; Gorczynski, R. J. *Life Sci.* **1982**, *31*, 899.
(2) Erhardt, P. W.; Woo, C. M.; Gorczynski, R. J.; Anderson, W. G. *J. Med. Chem.* **1982**, *25*, 1402.
(3) Erhardt, P. W.; Woo, C. M.; Anderson, W. G.; Gorczynski, R. J. *J. Med. Chem.* **1982**, *25*, 1408.
(4) Hoefle, M. L.; Hastings, S. G.; Meyer, R. F.; Corey, R. M.; Holmes, A.; Stratton, C. D. *J. Med. Chem.* **1975**, *18*, 148.
(5) Sakanashi, M. *Trends Pharmacol. Sci.* **1981**, *2*, 234.
(6) Maseri, A.; L'Abbate, A.; Baroldi, G.; Chierchia, S.; Marzilli, M.; Ballestra, A. M.; Severi, S.; Parodi, O.; Biagini, A.; Distante, A.; Pesola, A. *N. Engl. J. Med.* **1978**, *299*, 1271.
(7) Smith, L. H. *J. Appl. Chem. Biotechnol.* **1978**, *28*, 201.
(8) Large, M. S.; Smith, L. H. *J. Med. Chem.* **1980**, *23*, 112.
(9) Barlow, J. J.; Main, B. G.; Snow, H. M. *J. Med. Chem.* **1981**, *24*, 315.
(10) Hill, A. J.; Aspinall, S. R. *J. Am. Chem. Soc.* **1939**, *61*, 822.

0022-2623/83/1826-1109$01.50/0