Discriminant Analysis and Structure-Activity Relationships. 1. Naphthoquinones

Gopalakrishnan Prakash and Ernest M. Hodnett*

Department of Chemistry, Oklahoma State University, Stillwater, Oklahoma 74074. *Received March 7, 1977*

Discriminant analysis has been used to study the data for naphthoquinones as antitumor agents in three different animal tumor systems. In each case the most significant variables for classifying the compounds into two groups according to their antitumor activities were determined by a stepwise procedure. The usefulness of discriminant analysis in the design of drugs is discussed.

Discriminant analysis may be used to evaluate the abilities of variables to distinguish between two or more groups. The only requirement for applications to chemotherapy studies is that a series of compounds be graded into two or more classes by some biological property. Differences between two groups of compounds can usually be observed if one considers the average values of several variables for compounds of each group. However, some variables may be more significant than others in placing an individual compound in one group rather than in another. The function of discriminant analysis is to find a linear combination of factors that will best discriminate between the two groups, to weight the importance of these factors, and to indicate the positive or negative effect of these factors.

Discriminant analysis has been used very little in the study of drug design but can be a useful approach.[1] Martin et al.[2] have reported its use to study the relationship between structure and the inhibition of monoamine oxidase by aminotetralins and aminoindans. By the use of two variables, E_s^c and an indicator variable to represent the position of substitution, these authors were able to place 19 compounds out of 20 into the correct group (of two groups) on the basis of in vivo potency.

Discriminant analysis is particularly useful in the study of antitumor effects of synthetic compounds since dose–response curves are not available for many compounds, but some other indication of antitumor activity is. Antitumor results are often reported as T/C values for only one dose level, usually the optimum dose. T/C is the biological effect in animals treated with the drug compared to that in untreated animals. For solid tumors the parameter measured is the tumor weight inhibition after a given time and mode of treatment (TWI). For ascitic tumor systems the parameter measured is the increase in life span of the animals (ILS). Results of discriminant analysis can indicate which structural variations are most significant in distinguishing inactive and active compounds although the dose of each compound is not specified.

Further, the method is helpful in predicting the antitumor activity of an untested compound and therefore can speed drug design. Discriminant analysis provides a classification function for each group that indicates the relative importance of each variable in making that classification. On the basis of these functions, the probability of a new compound not in the original series belonging in a particular group can be calculated. The signs of the coefficients in the classification function indicate whether a particular parameter should be increased or decreased in order to modify the biological activity in the desired direction. Compounds with desirable values of the parameters may have their group classifications predicted, based on the original data set, before the compounds are prepared and tested in the laboratory.

We have found very useful the BMDO7M stepwise discriminant analysis program, revised December 1975, developed at the Health Sciences Computing Facility, University of California at Los Angeles.[3] This program performs a multiple discriminant analysis by selecting the independent variables one at a time in a stepwise manner to establish a classification function for each group. At each consecutive step the program chooses the variable with the greatest F value for entry into the classifications. Should the F value for a particular variable become too small as other variables are added, that variable is eliminated from the classification function. The specific procedures for using the BMDO7M program are documented in a clear and usable manner.[4] Some of the computational methods are summarized here.

The program first calculates and prints the overall mean for each variable, the group mean for each variable in each group, the standard deviation for each variable in each group, the within and total cross-product matrices, the within-group covariance matrix, and the within-groups correlation matrix.

In the first step of the stepwise procedure for selection of variables for the classification functions, the variable with the greatest partial F value is entered. This variable is the one that alone can best describe the difference between the groups. In each subsequent step, one variable is added or removed from the classification functions according to the following procedures.

(a) If one or more variables that have already been entered into the classification functions have an F value to remove that is smaller than the value set for the program (normally 0.005 unless modified by the user), then the variable with the smallest F value will be removed.

(b) If no variable is removed from the classification functions, then a variable not in classification functions is chosen by one of three alternative procedures. We have chosen to use throughout this work option (1) that selects the variable with the greatest F value to enter. This variable is the one that together with the variables already entered in the classification functions provides the best discrimination between the groups. At each step a classification function is derived for each group, the compounds are classified into each group based on the classification functions just derived, and the numbers of cases correctly and incorrectly classified are printed.

A classification function for each group k of n members is calculated as

$$c_{kj} = (n - g) \sum_{i=1}^{r} \bar{x}_{kj} a_{ij}$$

$$i = 1, 2, \ldots, r; k = 1, 2, \ldots, g$$

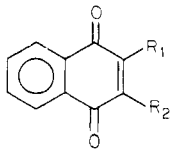
where c_{kj} is the coefficient for variable v_j of group k, g is the number of groups used for the analysis, r is the number of variables used, i and j indicate the variables, a_{ij} is the determinant of the inverse within-group cross-products matrix, and $\bar{x}_{kj}$ is the group mean for variable j. A constant term, c_{k0}, is calculated for each classification function.

$$c_{k0} = -1/2 \sum_{j=1}^{r} c_{kj} \bar{x}_{kj}$$

$$k = 1, 2, \ldots, g$$

0022-2623/78/1821-0369$01.00/0

Table I. L1210 Leukemia in CDBA and BDH1 Mice



R_1	R_2	$\pi(1)$	$F(1)$	$R(1)$	MR(1)	$\pi(2)$	$F(2)$	$R(2)$	MR(2)	$E_{1/2}$, V	ILS
Cl	H	+0.71	+0.41	−0.15	0.60	0.00	0.00	0.00	0.10	−0.141	0
OH	H	−0.67	+0.29	−0.64	0.28	0.00	0.00	0.00	0.10	−0.330	0
OCH_3	H	−0.02	+0.26	−0.51	0.79	0.00	0.00	0.00	0.10	−0.283	0
$OCOCH_3$	H	−0.64	+0.41	−0.07	1.25	0.00	0.00	0.00	0.10	−0.140	0
NH_2	H	−1.2	+0.02	−0.68	0.54	0.00	0.00	0.00	0.10	−0.372	0
NHC_6H_5	H	+1.37	−0.02	−0.38	3.00	0.00	0.00	0.00	0.10	−0.355	0
CH_3	CH_3	+0.56	−0.04	−0.13	0.56	+0.56	−0.04	−0.13	0.56	−0.292	0
CH_3	OCH_3	+0.56	−0.04	−0.13	0.56	−0.02	+0.26	−0.51	0.79	−0.268	0
OH	CH_3	−0.67	+0.29	−0.64	0.28	+0.56	−0.04	−0.13	0.56	−0.350	0
Br	Br	+0.86	+0.44	−0.17	0.89	+0.86	+0.44	−0.17	0.89	−0.167	0
Cl	$N(CH_3)_2$	+0.71	+0.41	−0.15	0.60	+0.18	+0.10	−0.92	1.56	−0.279	0
OCH_3	OCH_3	−0.02	+0.26	−0.51	0.79	−0.02	+0.26	−0.51	0.79	−0.274	0
H	H	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.10	−0.164	2
CH_3	H	+0.56	−0.04	−0.13	0.56	0.00	0.00	0.00	0.10	−0.224	14
SCH_3	H	+0.61	+0.20	−0.18	1.38	0.00	0.00	0.00	0.10	−0.283	4
Cl	Cl	+0.71	+0.41	−0.15	0.61	+0.71	+0.41	−0.15	0.60	−0.146	13
C_2H_5	H	+1.02	−0.05	−0.10	1.03	0.00	0.00	0.00	0.10		2
$COCH_3$	H	−0.55	+0.32	+0.20	1.12	0.00	0.00	0.00	0.10		0
SC_2H_5	H	+1.07	+0.23	−0.18	1.84	0.00	0.00	0.00	0.10		3
OH	$CH_2C_6H_5$	−0.67	+0.29	−0.64	0.28	+2.01	−0.08	−0.01	3.00		0
OH	$COCH_3$	−0.67	+0.29	−0.64	0.28	−0.55	+0.32	+0.20	1.12		0
CH_3	SCH_3	+0.56	−0.04	−0.13	0.56	+0.61	+0.20	−0.18	1.38		0
CH_3	SC_2H_5	+0.56	−0.04	−0.13	0.56	+1.07	+0.23	−0.18	1.84		0
OH	Br	−0.67	+0.29	−0.64	0.28	+0.86	+0.44	−0.17	0.89		2
Cl	$NHCH_3$	+0.71	+0.41	−0.15	0.60	−0.47	−0.11	−0.74	1.03		0
OH	Cl	−0.67	+0.29	−0.64	0.28	+0.71	+0.41	−0.15	0.60		0
OH	NH_2	−0.67	+0.20	−0.64	0.28	−1.2	±0.02	−0.68	0.54		11

For purposes of predicting the classification of each case (or compound), the mth classification function, s_{lmk}, can be evaluated for case k of group l

$$s_{lmk} = c_{m0} + \sum_{j=1}^{r} c_{mj} x_{lkj}$$

where c_{m0} is the constant term in the classification function for group m, c_{mj} is the coefficient for variable j in the classification function for group m, and x_{lkj} is the value of variable j for case k of group l. The posterior probability, P_{lmk}, of case k in group l having come from group m can be calculated

$$P_{lmk} = \frac{p_m \exp(s_{lmk})}{\sum_{i=1}^{g} p_i \exp(s_{lik})}$$

where p_m is the prior probability of group m, p_i is the prior probability of group i, and s_{lik} is the classification function evaluated for case k of group i. This program calculates the posterior probability of each compound (case) belonging to each group on the basis of the classification functions developed in the last step. The stepwise procedure terminates at a predetermined number of steps or when no variable meets the criteria (F value) set for entry or removal.

We have applied discriminant analysis to published data for the effectiveness of naphthoquinones against three different animal tumor systems. Our purpose was to determine the usefulness of this method of analysis in the design of more active compounds.

L1210 Leukemia in CDBA and BDH1 Mice. A search for naphthoquinones that are substituted in the quinone ring only, that have been tested against L1210 leukemia in mice, and that have known $E_{1/2}$ values yielded the first 16 naphthoquinones shown in Table I. Eleven more naphthoquinones (in the last part of Table I) were also used in this analysis although $E_{1/2}$ values were not available for them. The antitumor data are from the work of Driscoll et al.[5] Values of substituent constants π, F, R, and MR (scaled by 0.1) are listed by Hansch et al.[6] Values of $E_{1/2}$ for each compound are those given by Holmes.[7] Of these 27 naphthoquinones 19 are inactive against L1210 leukemia (group 1) and only eight are slightly active (ILS, 2% or more) (group 2).

Discriminant analysis of the 27 compounds in Table I indicates a high dependence on MR(2) ($F = 2.69 > F_{1,25,0.80} = 1.73$) and also MR(2) and $\pi(2)^2$ ($F = 2.15 > F_{2,24,0.80} = 1.72$). The latter pair of variables classified 19 of the 27 compounds correctly. However, this result is biased by the fact that for monosubstituted compounds the substituent is placed in position 2, leaving H as a substituent in position 3 for many cases. Since these two positions are identical for naphthoquinone, no valid reason (except for nomenclature purposes) exists for always placing a substituent in position 2.

Therefore, the substituents R_1 and R_2 were randomly distributed between positions 2 and 3, even for monosubstituted compounds. For one distribution, variable MR(2) has an extremely high significance in discriminating between the two groups ($F = 12.2 > F_{1,25,0.995} = 9.48$); MR(2) and MR(1)^2 also have a high confidence level ($F = 8.78 > F_{2,24,0.995} = 6.66$). However, when the substituents are again randomized between the two positions, the confidence levels are much lower; for MR(1), $F = 1.24 < F_{1,24,0.80} = 1.73$, and for MR(1) and $R(2)$, $F = 1.32 < F_{2,24,0.80} = 1.72$. Another arrangement of the groups gave for $R(2)$, $F = 1.58 < F_{1,25,0.80} = 1.73$, and for $R(2)$ and MR(1), $F = 1.31 < F_{2,24,0.80} = 1.72$. It is evident that since positions 2 and 3 in naphthoquinone are identical, the random

Table II. Significance of Combinations of Variables in Discriminating between Groups in the L1210 System

Variables	$F_{2,24}$	Confidence level, %	Variables	$F_{3,23}$	Confidence level, %	Variables	$F_{4,22}$	Confidence level, %
MR, π^2	1.89	80	MR, π, π^2	2.26	90	MR, π, π^2, R	1.67	80
MR, π	1.84	80	MR^2, π, π^2	2.05	80	MR, π, MR^2, π^2	1.63	80
MR^2, π^2	1.81	80	MR, R, π^2	1.59	<80	MR, F, π, π^2	1.62	~80
MR^2, π	1.59	<80	MR^2, R, π^2	1.56	<80	MR^2, π, π^2, R	1.54	~80

distribution of groups cannot give consistent, significant results for individual substituent parameters.

Consequently, the sums of the parameters for substituents R_1 and R_2, in position 2 and 3, π, F, R, MR, MR^2, and π^2, have been used for further discriminant analysis of the naphthoquinones. The most significant combinations of variables for this case are given in Table II with the corresponding F values for discrimination between the two groups. Although several combinations of variables give similar discrimination between the two groups, the variables π, MR, and π^2 seem to be considerably better than other combinations based on their F values. These three variables allow the classification of 17 of the 19 inactive compounds and five of the eight slightly active compounds. Discrimination between the two groups is at the 80% confidence level ($F = 2.26 > F_{3,23,0.80} = 1.68$).

The two groups may be described by the classification functions below that utilize all six variables.

inactive compounds (group 1)

$$-1.30\,\pi + 19.1\,F + 19.9\,R + 19.2\,\mathrm{MR} - 4.50\,\mathrm{MR}^2 - 0.06\,\pi^2 - 8.61$$

slightly active compounds (group 2)

$$-0.28\,\pi + 16.4\,F + 17.4\,R + 17.4\,\mathrm{MR} - 4.60\,\mathrm{MR}^2 + 0.96\,\pi^2 - 8.02$$

A single function, obtained by subtracting the classification function for group 1 from that for group 2, may be used to describe the difference in the two groups (the $F_{1,19}$ to remove is given in parentheses).

$$+1.02\ (1.49)\,\pi - 2.7\ (0.041)\,F - 2.5\ (0.041)\,R - 1.8\ (0.35)\,\mathrm{MR} - 0.1\ (0.015)\,\mathrm{MR}^2 + 1.02\ (2.27)\,\pi^2 + 0.59$$

Some variables in this discriminant function are not highly significant. The most significant variables, π, π^2, and MR, have the largest F values (in parentheses). The less significant variables do not increase the ability of the function to discriminate between the groups. The complete function predicts correctly 22 of the 27 compounds at a confidence level of much less than 80% ($F = 0.83 \ll F_{7,19,0.80} = 1.59$), whereas the three most significant variables predict correctly 22 of the 27 compounds at a confidence level of better than 80% ($F = 2.26 > F_{3,23,0.80} = 1.68$). We use the more complete functions to show the relative importance of all the variables in discriminating between the groups based on their antitumor activities.

The sign for each term in the discriminant function indicates changes needed to convert compounds in group 1 to group 2 or to increase the activity of the compounds in controlling L1210 leukemia. Therefore, for greater activity in this system, compounds with large positive values of π and π^2 and smaller values of MR are needed. The mean values of π and π^2 for compounds in group 2 are greater than those for group 1, and the mean values of MR are smaller for group 2 than for group 1.

Discriminant analysis of the data in the L1210 system for the 16 compounds for which $E_{1/2}$ values could be found developed the following classification functions for the two groups ($F_{1,11}$ values are in parentheses).

inactive compounds

$$-2.26\ (1.83)\,\pi - 3.33\ (1.67)\,R + 15.0\ (0.35)\,\mathrm{MR} - 3.51\ (0.02)\,\mathrm{MR}^2 - 6.87$$

slightly active compounds

$$-0.35\ (1.83)\,\pi + 0.33\ (1.67)\,R + 12.6\ (0.35)\,\mathrm{MR} - 3.35\ (0.02)\,\mathrm{MR}^2 - 5.08$$

Variables F, π^2, and $E_{1/2}$ do not appear due to their relatively low significance. These classification functions provide discrimination between the two groups at the 80% confidence level ($F = 1.84 > F_{4,11,0.80} = 1.80$). Use of R alone allows for better discrimination ($F = 5.48 > F_{1,14,0.95} = 4.60$). This variable classifies 11 of the 12 inactive compounds correctly and one of the four slightly active compounds.

Solid Sarcoma 180 of Swiss Mice. Antitumor effects against solid sarcoma 180 of Swiss mice[5] and the half-wave potentials[7] for 13 naphthoquinones are listed in the first part of Table III. The last eight naphthoquinones in Table III have no published values of $E_{1/2}$. Eleven of these 21 naphthoquinones showed inhibitions of tumor weight gains in this tumor system of less than 20% and are classed as inactive (group 1); ten produced tumor weight inhibitions of 20–50% and are considered moderately active (group 2). Values of the substituent constants π, F, R, and MR are those listed by Hansch[6] and are summed for both groups or atoms, R_1 and R_2, since again these positions are identical. When all 21 compounds are subjected to discriminant analysis, a single discriminant function is generated ($F_{1,13}$ values to remove each variable are in parentheses).

$$-2.81\ (2.80)\,\pi + 172\ (2.78)\,F + 191\ (3.04)\,R + 12.4\ (7.07)\,\mathrm{MR} - 1.36\ (1.23)\,\mathrm{MR}^2 - 3.5\ (4.72)\,\pi^2 - 6.9$$

Use of this discriminant function allows the correct classification of 19 of the 21 compounds at the 95% confidence level ($F_{7,13} = 3.69$).

However, use of three variables, R, MR, and π^2, provides correct classification of 18 of the 21 compounds at the 99% confidence level ($F = 5.29 > F_{4,16,0.99} = 5.18$). Other combinations of three and four variables that can discriminate well between the two groups of compounds are shown in Table IV. The two classification functions involving R, MR, and π^2 are ($F_{1,17}$ values are in parentheses)

inactive compounds

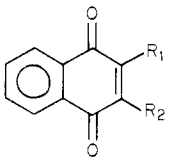
$$-7.56\ (4.15)\,R - 0.03\ (15.4)\,\mathrm{MR} + 2.28\ (6.85)\,\pi^2 - 3.61$$

moderately active compounds

$$-2.83\ (4.15)\,R + 4.05\ (15.4)\,\mathrm{MR} - 0.62\ (6.85)\,\pi^2 - 4.98$$

The coefficients for the moderately active compounds are for π^2 more negative and for R and MR more positive than

Table III. Solid Sarcoma 180 in Swiss Mice



R_1	R_2	π	F	R	MR	$E_{1/2}$, V	TWI, %
H	H	0.00	0.00	0.00	0.20	−0.164	5
OH	H	−0.67	+0.29	−0.64	0.39	−0.330	0
OCH_3	H	−0.02	+0.26	−0.51	0.89	−0.283	0
CH_3	CH_3	+1.12	−0.08	−0.26	1.13	−0.292	6
CH_3	OCH_3	+0.54	+0.22	−0.64	1.35	−0.268	0
OH	CH_3	−0.11	+0.25	−0.77	0.85	−0.350	10
Cl	Cl	+1.42	+0.82	−0.30	1.20	−0.145	0
NH_2	H	−1.23	+0.02	−0.68	0.64	−0.372	17
$NHCOCH_3$	H	−0.97	+0.28	−0.26	1.60	−0.224	39
NHC_6H_5	H	+1.37	−0.02	−0.38	3.11	−0.355	22
SCH_3	H	+0.61	−0.20	−0.18	1.48	−0.283	29
Cl	$N(CH_3)_2$	+0.89	+0.51	−1.07	2.16	−0.279	16
OCH_3	OCH_3	−0.04	+0.52	−1.02	1.57	−0.274	44
C_2H_5	H	+1.02	−0.05	−0.10	1.13		24
SC_2H_5	H	+1.07	+0.23	−0.18	1.94		20
OH	$CH_2C_6H_5$	+1.34	+0.21	−0.65	3.29		39
CH_3	SCH_3	+1.17	+0.16	−0.31	1.95		43
CH_3	SC_2H_5	+1.63	+0.19	−0.31	2.41		0
OH	Br	+0.19	+0.73	−0.81	1.17		32
Cl	$NHCH_3$	+0.24	+0.30	−0.89	1.64		40
OH	Cl	+0.04	+0.70	−0.79	0.89		0

Table IV. Significance of Combinations of Variables in Discriminating between Groups in the Solid Sarcoma 180 System

Variables	$F_{3,17}$	Confidence level, %	Variables	$F_{4,16}$	Confidence level, %
R, MR, π^2	5.29	99	π, R, MR, π^2	4.57	97.5
MR, MR^2, π^2	3.54	95	R, MR, MR^2, π^2	3.87	97.5
F, MR, π^2	3.31	95	F, R, MR, π^2	3.81	97.5
π, MR, π^2	3.28	95	π, R, MR, MR^2	3.13	95

Table V. Correlation Matrix for the Parameters Used in the Solid Sarcoma 180 Tumor System

Variable	π	F	R	MR	MR^2	π^2
π	1.000					
F	−0.044	1.000				
R	0.276	−0.514	1.000			
MR	0.652	−0.044	−0.137	1.000		
MR^2	0.581	−0.142	−0.088	0.957	1.000	
π^2	0.582	−0.230	0.372	0.658	0.640	1.000

those for the inactive group. Therefore, substituents with small π^2 values, positive R values, and large MR values should have greater antitumor activities in this system. The within-groups correlation matrix given in Table V shows high correlation only between MR and MR^2 (the means of these variables for all compounds are 1.27 and 2.17).

When the data for the 13 compounds having known $E_{1/2}$ values are analyzed, a discriminant function containing most of the variables is obtained ($F_{1,5}$ values are in parentheses).

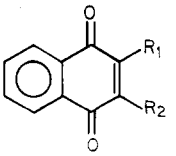
$$-12.9\ (9.43)\ \pi + 674\ (6.80)\ F + 7.34\ (7.03)\ R + 28.5\ (33.8)\ \mathrm{MR} - 6.93\ (2.04)\ \pi^2 - 35\ (0.20)\ E_{1/2} - 34.6$$

The significance of the variable MR was too small to be considered. The discrimination between the groups is high ($F = 6.43 > F_{7,5,0.95} = 4.88$), and all 13 compounds are classified correctly. When only MR and π^2 are used, the discrimination between the groups is at the 97.5% confidence level ($F = 4.99 > F_{2,10,0.975} = 4.83$), but one compound in each group is classified incorrectly. $E_{1/2}$ does not appear to be a significant variable in discriminating between the two groups according to their activity in this system.

Ascitic Sarcoma 180 in Swiss Mice. Lin and Sartorelli have synthesized and tested for antitumor activities a variety of benzoquinones and naphthoquinones.[8-12] These compounds are thought to undergo bioreduction in vivo to form *o*-quinone methides that alkylate DNA and other vital cellular components.[13] Recently, the same authors have synthesized five benzoquinones and 20 naphthoquinones, measured their polarographic half-wave potentials, and determined their effect on survival times in Swiss mice with ascitic sarcoma 180.[14] The results qualitatively indicate a correlation between the half-wave potentials of the quinones and their activities as antitumor agents. In an effort to quantify the findings of these authors, we have applied discriminant analysis to some of the naphthoquinones that they reported.

Fifteen naphthoquinones for which $E_{1/2}$ values were reported by Lin and Sartorelli[14] are listed in Table VI, together with the largest T/C values that Sartorelli and co-workers have published for these compounds. Four compounds having substituents in the aromatic ring of naphthoquinone are omitted from this analysis because of their small number. These 15 compounds were divided into two groups according to their T/C values. Group 1 contained eight moderately active compounds and group 2, seven active compounds. The substituent parameters are those reported by Hansch et al.[6] except that values of F and R have been estimated for compounds 11–13 and 15. These substituent values have been summed for two positions on the quinone ring, since each substituent is adjacent to one of the carbonyl atoms. Values of MR are scaled by a factor of 0.1 in order to bring their magnitudes closer to those of the other substituent parameters.

Table VI. Ascitic Sarcoma 180 in Mice



Compd	R_1	R_2	$E_{1/2}$, V	π	F	R	MR	T/C
1	CH_2Cl	CH_3	−0.32	+0.73	+0.06	−0.10	1.61	1.70
2	CH_2Cl	H	−0.24	+0.17	+0.10	+0.03	1.15	2.07
3	CH_2Br	H	−0.23	+0.79	+0.10	+0.05	1.44	2.18
4	CH_2Br	CH_3	−0.31	+1.35	+0.06	−0.08	1.90	2.21
5	CH_2Cl	CH_2Cl	−0.29	+0.34	+0.20	+0.06	2.10	1.63
6	CH_2Br	CH_2Br	−0.27	+1.58	+0.20	+0.10	2.68	2.29
7	CH_2Cl	C_6H_5	−0.28	+2.13	+0.18	−0.05	3.58	1.86
8	CH_2Cl	SC_2H_5	−0.27	+1.24	+0.33	−0.15	2.89	2.40
9	CH_2Br	Br	−0.24	+1.65	+0.54	−0.12	2.23	2.30
10	CH_2Br	Cl	−0.25	+1.50	+0.51	−0.10	1.94	2.30
11	CH_2OAc	H	−0.22	−0.17	0.00	0.00	1.75	1.48
12	CH_2OAc	CH_3	−0.28	+0.39	−0.04	−0.13	2.21	2.06
13	CH_2OAc	CH_2OAc	−0.25	−0.34	0.00	0.00	3.30	2.02
14	CH_2Cl	$NHCOC_6H_5$	−0.23	+0.66	+0.19	−0.24	4.51	1.1
15	CH_2Cl	SC_6H_5	−0.23	+2.49	+0.33	−0.15	4.48	1.8

The complete discriminant function for these two groups is ($F_{1,6}$ to remove in parentheses)

$$14.4\ (2.63)\ \pi - 14.9\ (0.078)\ F - 7.8\ (0.020)\ R + 1.9\ (0.041)\ \mathrm{MR} - 0.75\ (0.21)\ \mathrm{MR}^2 - 5.48\ (1.70)\ \pi^2 + 8\ (1.59)\ E_{1/2} + 15.8$$

These variables can classify all of the moderately active compounds correctly and six of the seven active compounds, although the confidence level is less than 80% ($F = 1.73 < F_{8,6,0.80} = 2.04$). However, four variables, π, MR, π^2, and $E_{1/2}$, can place all compounds into their proper groups and can discriminate between the groups at a confidence level of more than 95% ($F = 4.38 > F_{4,10,0.95} = 3.48$). The classification functions are ($F_{1,10}$ to remove in parentheses)

moderately active compounds

$$-28.0\ (7.60)\ \pi + 9.81\ (5.30)\ \mathrm{MR} + 9.44\ (3.70)\ \pi^2 - 455\ (3.57)\ E_{1/2} - 71.1$$

active compounds

$$-18.6\ (7.60)\ \pi + 7.30\ (5.30)\ \mathrm{MR} + 6.30\ (3.70)\ \pi^2 - 392\ (3.57)\ E_{1/2} - 52.8$$

The parameter $E_{1/2}$ alone cannot distinguish between the two groups at any reasonable level of significance ($F = 0.0057 \ll F_{1,13,0.80} = 1.82$). $E_{1/2}$ has low correlation with all other parameters.

Substituents with small positive values of π and small values of MR should be given consideration for further study.

Discussion of Results

Activities against the L1210 system of the first 16 compounds in Table I have low dependence on the half-wave potential, $E_{1/2}$. When all 27 compounds in Table I are used (without $E_{1/2}$ as a variable), the best combination of variables is MR, π, and π^2 at a confidence level of 80%. We believe that this low significance can be attributed to the fact that by the usual criterion none of the naphthoquinones in Table I are significantly active in the L1210 tumor system. The arbitrary classification of a compound with an ILS of 2% as a slightly active compound is not realistic. If the customary definition of active compounds in this system (ILS of 25%) is used, only one naphthoquinone would be classed as active—the sodium salt of lapachol. It is interesting to note that the three most significant variables, MR, π, and π^2, classify 20 of the 27 compounds in Table I as inactive. An increase in activity against the L1210 tumor system in the 1,4-naphthoquinones may be afforded by substituents having small values of MR and large positive values of π.

In the solid sarcoma 180 system excellent dependence on MR, R, and π^2 at the 99% confidence level was demonstrated. Although MR shows a high correlation with MR^2, none of the other variables are collinear. The classification functions indicate that substituents with large values of MR, small values of π, and positive values of R are desirable for maximum activity in this system.

Discriminant analysis of the results with the ascitic sarcoma 180 showed that the four variables, π, π^2, MR, and $E_{1/2}$, were able to discriminate between the two groups at the 95% confidence level. Substituents with small positive values of π and small values of MR should improve the antitumor activity in this system.

Acknowledgment. We are grateful to Dr. William J. Dunn for his suggestion concerning the usefulness of discriminant analysis in structure–activity relationships. Dr. Robert Morrison and Dr. Don Holbert have given us valuable assistance in the application and interpretation of statistical methods to biological activities. The technical assistance of Al Bounds is gratefully acknowledged. We appreciate the helpful comments of two reviewers.

References and Notes

(1) Y. C. Martin, *Cancer Chemother. Rep., Part 2*, 4, 35 (1974).
(2) Y. C. Martin, J. B. Holland, C. H. Jarboe, and N. Plotnikoff, *J. Med. Chem.*, 17, 409 (1974).
(3) The Health Sciences Computing Facility is sponsored by NIH Special Research Resources Grant RR-3.
(4) W. J. Dixon, "BMD—Biomedical Computer Programs", University of California Press, Berkeley, Calif., 1971, p 214a.
(5) J. S. Driscoll, G. F. Hazard, H. B. Wood, and A. Goldin, *Cancer Chemother. Rep., Part 2*, 4, 1 (1974).
(6) C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nikiaitani, and E. J. Lien, *J. Med. Chem.*, 16, 1209 (1973).
(7) H. L. Holmes, "Structure–Activity Relationships", Vol. 2, Defence Research Establishment, Suffield, Ralston, Alberta, Canada, 1975, p 1229.
(8) A. J. Lin, L. A. Cosby, C. W. Shansky, and A. C. Sartorelli, *J. Med. Chem.*, 15, 1247 (1972).
(9) A. J. Lin, R. S. Pardini, L. A. Cosby, B. J. Lillis, C. W.

Shansky, and A. C. Sartorelli, J. Med. Chem., 16, 1268 (1973).
(10) A. J. Lin, C. W. Shansky, and A. C. Sartorelli, J. Med. Chem., 17, 558 (1974).
(11) A. J. Lin, R. S. Pardini, B. J. Lillis, and A. C. Sartorelli, J. Med. Chem., 17, 668 (1974).
(12) A. J. Lin, B. J. Lillis, and A. C. Sartorelli, J. Med. Chem., 18, 917 (1975).
(13) A. J. Lin, L. A. Cosby, and A. C. Sartorelli, Cancer Chemother. Rep., Part 2, 4, 23 (1974).
(14) A. J. Lin and A. C. Sartorelli, Biochem. Pharmacol., 25, 206 (1976).

Activity–Electroreduction Relationship of Antimicrobial Metronidazole Analogues

Yie W. Chien* and Seth S. Mizuba

Searle Laboratories, Skokie, Illinois 60076. *Received July 28, 1977*

The antimicrobial action of metronidazole against anaerobes is thought to be by activation due to its competition of electrons with ferredoxins, an electron-transfer protein. In this investigation, the electroreduction properties of metronidazole and eight analogues were studied by a sensitive ac polarographic technique in comparison with clostridial ferredoxin. The results showed that all the metronidazole analogues had an ac reduction peak potential that was 44–122 mV less negative than that for clostridial ferredoxin. Using *Clostridium pasteurianum* and *Trichomonas vaginalis* as the test microorganisms, the antimicrobial activities of these metronidazole analogues were determined. A theoretical expression was derived to define the relationship between the antimicrobial activity of metronidazole analogues and their activation free energy for electroreduction and lipophilicity for cell permeation. Statistical analyses of the experimental data suggested that the growth inhibition of metronidazole analogues toward *Cl. pasteurianum* depended on their activation free energy. For the growth inhibition of *T. vaginalis*, the lipophilicity of metronidazole analogues was as important as the activation free energy, as expected from the theoretical model. The competitive electroreduction between metronidazole analogues and ferredoxin was also examined. The addition of various concentrations of a metronidazole analogue to a ferredoxin solution had no effect, except to reduce the electroreductivity of ferredoxin's S–Fe bondings. This effect was observed to be directly proportional to the drug concentrations added. Thus, it was concluded that an active metronidazole analogue requires an electroreduction potential less negative than ferredoxin to be a better electron acceptor and a lower activation free energy of proton transfer to be irreversibly reduced itself to a polar derivative. This reduced species may subsequently interfere with the metabolic activity of the anaerobes, thus eliciting its antimicrobial activity.

Metronidazole [1-(2-hydroxyethyl)-2-methyl-5-nitroimidazole][1] and other nitroimidazole derivatives are used extensively to treat infections caused by anaerobic protozoa and bacteria.[2] These drugs were found to be selectively absorbed and had a cytotoxic action on anaerobes.[3-5] Their toxicity for aerobic microorganisms and for mammalian cells is low.[6,7]

Recently, Edwards and his co-workers[3] investigated the electroreduction of some antimicrobial agents and observed that the active ones, e.g., 5-nitroimidazoles, have a less negative redox potential, while the inactive ones, e.g., 4-nitroimidazoles and 4-nitropyrazole, have a more negative redox potential than that (−0.470 V) of ferredoxin, an electron-transfer protein required in the pyruvate phosphoroclastic system of anaerobes. This observation led them to conclude that the active 5-nitroimidazole acts as an efficient electron sink by accepting electrons from a reduced ferredoxin molecule via its nitro ($-NO_2$) group; the nitro group becomes irreversibly reduced in the process, and the reduction products bind to nucleic acids and inhibit the metabolic functions of anaerobes. Earlier studies by Miller et al.[8] and Howes with his associates[9] also suggested the importance of the electronegativity of the N_1 substituents of 5-nitroimidazoles in determining both their electron acceptability and their antimicrobial activity.

It has been well documented that the ferredoxin serves as an oxidation–reduction enzyme in anaerobes, transferring electrons from a low potential donor to electron-accepting biochemicals.[10] The interruption of this vital pyruvate phosphoroclastic reaction may well be related to the selective toxicity observed in the anaerobes.[2-5]

Recent analyses of the electroreduction characteristics of ferredoxins, using a sensitive alternating current (ac) polarographic technique, demonstrated that the electron-transport mechanisms of the ferredoxin molecules were linked closely to their sulfur–iron bondings.[11] The dissociation of the sulfur–iron bonds resulted in the formation of a free cysteinic SH group and the interruption of the electroactivity of ferredoxins. The advantages of using ac polarography rather than dc polarography for the mechanistical analysis of electrochemical phenomenon were also illustrated.[11,12]

In this investigation, the antimicrobial activity and the electroactivity of metronidazole and its eight analogues, with a variety of N_1 side chains, are analyzed to gain, hopefully, a better understanding of the mechanisms of their antimicrobial action.

Experimental Section

(A) Alternating Current Polarography of Metronidazole Analogues. Basically, the same experimental procedure reported earlier[11] was used in this investigation. Except where specified, a metronidazole analogue was used as obtained and a solution of 8×10^{-5} M was prepared in pH 6.01 McIlvaine buffer just prior to a polarographic measurement. For easy comparison, the ac polarographic analyses were performed on the same PAR electrochemistry system,[13] under similar conditions as those specified previously for ferredoxins.[11]

(B) Partition Studies. Drug solutions with a concentration of 8×10^{-5} M were freshly prepared in 1-octanol–saturated McIlvaine buffer (0.1 M, pH 6.01). Ten milliliters were vortexed and equilibrated with 10 mL of McIlvaine buffer-saturated 1-octanol. The drug concentration in the McIlvaine buffer phase before and after partitioning was measured spectrophotometrically and then utilized to calculate the magnitude of the partition coefficient.[14] The partition coefficient, PC, is defined as

$$PC = \frac{\text{equilibrium drug concn in 1-octanol phase}}{\text{equilibrium drug concn in McIlvaine buffer phase}} \quad (1)$$

0022-2623/78/1821-0374$01.00/0