predictive ability was not assessed on truly unknown compounds, a statistical procedure that produces unbiased estimates of predictive ability was used. Thus, it has been demonstrated that pattern-recognition procedures may have utility for the prediction of carcinogenic activity of compounds.

Acknowledgment. This research was sponsored by the National Cancer Institute through Contract N01 CP 75926. The computer used for this work was purchased with partial financial support of the National Science Foundation.

References and Notes

(1) B. A. Bridges, *Nature (London)*, **261**, 195 (1976).
(2) I. F. H. Purchase, E. Longstaff, J. Ashby, J. A. Styles, D. Anderson, P. A. Lefevre, and F. R. Westwood, *Nature (London)*, **264**, 624 (1976).
(3) E. J. Ariens, *Drug Des., 1971–1978*, 1–8 (1971–1978).
(4) A. Burger, "Medicinal Chemistry", Part I, Wiley-Interscience, New York, 1970.
(5) B. Bloom and G. E. Ullyot, Eds., "Drug Discovery", American Chemical Society, Washington, D.C., 1971.
(6) Wade Van Valkenburg, Ed., "Biological Correlations—The Hansch Approach", American Chemical Society, Washington, D.C., 1972, p 252.
(7) W. P. Purcell, G. E. Bass, and J. M. Clayton, "Strategy of Drug Design", Wiley-Interscience, New York, 1973.
(8) Y. C. Martin, "Quantitative Drug Design", Marcel Dekker, New York, 1978.
(9) Science Information Services Department, Franklin Institute Research Laboratories, "Structure Activity Correlation Bibliography: With Subject and Author Index", PB-240 658/5 GA, Mar 1975.
(10) W. J. Dunn, *Annu. Rep. Med. Chem.*, **8**, 313 (1973).
(11) R. D. Cramer, *Annu. Rep. Med. Chem.*, **11**, 301 (1976).
(12) C. Hansch, in "Advances in Linear Free Energy Relationships", Vol. 2, N. R. Chapman and J. Shorter, Eds., Plenum Press, New York, in press.
(13) P. N. Craig, in ref 6, p 115.
(14) W. G. Richards and M. E. Black, *Prog. Med. Chem.*, **11**, 67 (1975).
(15) R. E. Christoffersen, in "Quantum Mechanics of Molecular Conformations", B. Pullman, Ed., Wiley, New York, 1976.
(16) G. L. Kirschner and B. R. Kowalski, *Drug Des., 1978*, **8**, (1978).
(17) N. J. Nilsson, "Learning Machines", McGraw-Hill, New York, 1965.
(18) E. A. Patrick, "Fundamentals of Pattern Recognition", Prentice-Hall, Englewood Cliffs, N.J., 1972.
(19) H. C. Andrews, "Introduction to Mathematical Techniques in Pattern Recognition", Wiley-Interscience, New York, 1972.
(20) J. T. Tou and R. C. Gonzalez, "Pattern Recognition Principles", Addison-Wesley, Reading, Mass., 1974.
(21) V. L. Tal'roze, V. V. Raznikov, and G. D. Tantsyrev, *Dokl. Akad. Nauk SSSR*, **159**(1), 182 (1964).
(22) P. C. Jurs, B. R. Kowalski, and T. L. Isenhour, *Anal. Chem.*, **41**, 21 (1969).
(23) P. C. Jurs and T. L. Isenhour, "Chemical Applications of Pattern Recognition", Wiley-Interscience, New York, 1975.
(24) B. R. Kowalski, *Anal. Chem.*, **47**, 1152A (1975).
(25) M. L. McConnell, G. Rhodes, U. Watson, and M. Novotny, *J. Chromatogr.*, in press.
(26) J. S. Wishnok and M. C. Archer, *Br. J. Cancer*, **33**, 307 (1976).
(27) G. M. Singer, H. W. Taylor, and W. Lijinsky, *Chem.-Biol. Interact.*, **19**, 133 (1977).
(28) I. A. Smith, G. D. Berger, P. G. Seybold, and M. P. Serve, *Cancer Res.*, **38**, 2968 (1978).
(29) J. S. Wishnok, M. C. Archer, A. S. Edelman, and W. M. Rand, *Chem.-Biol. Interact.*, **20**, 43 (1978).
(30) L. B. Kier, R. J. Simons, and L. H. Hall, *J. Pharm. Sci.*, **67**, 725 (1978).
(31) A. J. Hopfinger and G. Klopman, *Chem.-Biol. Interact.*, in press.
(32) W. J. Dunn III and S. Wold, *J. Med. Chem.*, **21**, 1001 (1978).
(33) A. J. Stuper, W. E. Brugger, and P. C. Jurs, in "Chemometrics: Theory and Application", B. R. Kowalski, Ed., American Chemical Society, Washington, D.C., 1977, p 165.
(34) J. McCann, E. Choi, E. Yamasaki, and B. N. Ames, *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 5135 (1975).
(35) L. B. Kier and L. H. Hall, "Molecular Connectivity in Chemistry and Drug Research", Academic Press, New York, 1976.
(36) W. J. Murray, *J. Pharm. Sci.*, **66**, 1352 (1977).

Multivariate Analysis and Quantitative Structure–Activity Relationships. Inhibition of Dihydrofolate Reductase and Thymidylate Synthetase by Quinazolines

Bor-Kuan Chen, Csaba Horváth,*

Chemical Engineering Group, Department of Engineering and Applied Science, Yale University, New Haven, Connecticut 06520

and Joseph R. Bertino

Department of Pharmacology, School of Medicine, Yale University, New Haven, Connecticut 06510.
Received November 27, 1978

Quantitative structure–activity relationships (QSAR) have been established for the inhibition of dihydrofolate reductase and thymidylate synthetase by 2,4-diaminoquinazoline–glutamic acid analogues. For dihydrofolate reductase from both human acute lymphocytic leukemia cells and murine L1210R cells, QSAR's obtained with 50 quinazolines were similar. On the other hand, for the inhibition of thymidylate synthetase from murine L1210S cells and from *Lactobacillus casei*, QSAR's formulated on the basis of data measured with 33 compounds were different, indicating that the two enzymes are dissimilar. The use of multivariate statistics including cluster analysis, factor analysis, and discriminant analysis is shown to facilitate the formulation of a satisfactory correlation equation. The procedure is demonstrated by the development of QSAR for the inhibition of thymidylate synthetase.

Folate antagonists continue to be useful drugs in the treatment of certain neoplastic diseases. A great number of compounds have been synthetized and tested for biological activity in order to find more potent and less toxic anticancer agents. All clinically useful folate antagonists are inhibitors of the enzyme dihydrofolate reductase[1] (EC 1.5.1.3), and the screening of potential drugs usually commences with the measurements of the inhibitory

0022-2623/79/1822-0483$01.00/0

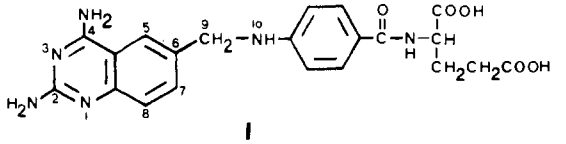
properties, I_{50}, of the substance. This approach goes back to the elegant studies by Baker and co-workers,[2] which has led to the design of several extremely potent enzyme inhibitors. The extensive data obtained by Baker and co-workers on the inhibition of dihydrofolate reductase have been analyzed by Hansch and his collaborators to find correlation between structure and activity.[3]

In the search for other enzymes of similar physiological role, interest has been focused on thymidylate synthetase (EC 2.1.1.45) which catalyzes the formation of thymidylate from 2-deoxyuridylate and, in the course of the reaction, 5,10-methylenetetrahydrofolate, a cofactor of the enzyme, is converted into dihydrofolate.[4] This enzyme is a target for antitumor agents since it plays an important role in DNA synthesis as does dihydrofolate reductase.[5] The enzyme has recently been isolated from *Lactobacillus casei* and mouse leukemia cell line L1210.[6] In the present study, the inhibition of the two enzymes by a certain group of quinazoline compounds is analyzed in order to shed light on the relationship between their chemical structure and inhibitory activity.

When sufficient quantitative biological data are available for a series of congeners, then the semiempirical linear free energy related method of Hansch[7] is used most frequently to establish quantitative structure–activity relationships (QSAR). This approach employs certain physicochemical parameters characteristic for the substituents and seeks the best statistical correlation between the biological activities of congeners and the physicochemical parameters of the substituents. The correlation technique has been expanded[3b] by the incorporation of indicator variables that are related to "compound identification" used in the Free–Wilson method.[8] In our study, the expanded method is employed to establish the QSAR of certain 2,4-diaminoquinazoline–glutamic acid analogues with regard to the inhibition of dihydrofolate reductase and thymidylate synthetase. We found that the use of various multivariate statistical methods, such as discriminant analysis,[9,10] factor analysis,[11,12] and cluster analysis,[13] can greatly facilitate the finding of the appropriate correlation equation for QSAR. This approach will be illustrated in the case of thymidylate synthetase inhibition.

Structural Parameters of Quinazolines

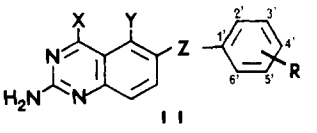
Congeners of *N*-[*p*-[[(2,4-diamino-6-quinazolinyl)-methyl]amino]benzoyl]glutamic acid (I) were used to



I

inhibit dihydrofolate reductase and thymidylate synthetase from different sources.

All compounds can be derived from the general formula (II) and differ only in the chemical nature of the sub-



II

stituents X, Y, Z, and R.

Appropriate values of the hydrophobicity, π, molar refractivity, MR, and Hammett constant, σ, for the substituents were taken from Leo et al.[14a] and Hansch et al.[15] or estimated following Hansch's approach.[14b] MR values were scaled by 0.1 to make them commensurable with the other parameters.[3b] Altogether, seven variables, I-1 to I-7, were used. For the model compound (I), each indicator variable is given the value 0. I-1 takes the value of 1 for 4-OH or 4-SH compounds, and I-2 is given the value of 1 when the substituent in the position 5 is other than hydrogen. I-3 is equal to 1 for the derivatives with Z = $-NHCH_2-$, and I-4 assumes the value of 1 for congeners having Z groups other than $-CH_2NH-$ or $-NHCH_2-$. I-5 is taken as 1 when the R group is not an amino acid derivative. If the amino acid is not L-glutamic acid, then I-6 equals 1, and a value of 1 for I-7 indicates that R is not in the para but in another position of the phenyl ring. As shown later, some of the indicator variables will be used to express the contribution of certain substituents whose physicochemical parameters showed poor variation and high correlation.

QSAR for Inhibition of Dihydrofolate Reductase

Hynes et al.[16] have investigated the inhibition of dihydrofolate reductase (DHFR) from rat liver and bacteria by various quinazolines. The data were analyzed and QSAR has been formulated by Hansch and co-workers.[17]

In the present investigation, I_{50} values obtained by Bertino et al.[18] with quinazolines for dihydrofolate reductase from human acute lymphocytic leukemia (ALL) and mice L1210R leukemic cells have been used to study QSAR. The experimental data together with the predicted inhibition potency, which is calculated from the corresponding QSAR as shown below, are listed in Table I.

There are only three kinds of substituents (H, CH_3, and Cl) in position 5, and the π values for Cl and CH_3 are so close (0.56 and 0.71) that the use of indicator variable I-2 is preferable to that of physicochemical parameters. Because of the poor variation between the σ values, it was appropriate to use the corresponding indicator variables. In calculating the physicochemical parameters for the substituent in the 6 position, the π-R and MR-R values of the R group were treated separately.

Employing the approach of Hansch et al.[17] and using the leaps and bounds algorithm of Furnival and Wilson for best subset regression,[19] we obtained the following QSAR for the inhibition of DHFR from human leukemic cells

$$\log (1/I_{50}) = 10.12 - 2.87(I\text{-}1) + 0.29(I\text{-}2) - 0.38(\text{MR-6}) - 0.29(\pi\text{-R}) - 0.19(\text{MR-R}) \quad (1)$$

SE	0.45	0.16	0.14		
t stat	22.2	−17.4	2.1		
				0.11	0.06	0.07
				−3.5	−4.6	−2.8

$n = 47$; $r = 0.956$; $s = 0.42$; $F_{5,41} = 86.35$

where n refers to the number of data points used in the regression, r is the multiple correlation coefficient, and s is the standard error of estimation. All terms are significant based on two-tailed test of significance. As seen, the F value of 86.35 is much larger than the critical value at the 0.1% significance level ($F_{5,41,\alpha=0.001} = 5.13$) so that eq 1 is highly statistically significant. The only reason for deleting compound **9** in the regression analysis is that it has the largest standardized residual. The interdependence of the parameters π-R and MR-R is poor, as illustrated in Figure 1, despite the fact that $r = -0.70$. Therefore, both parameters are used in regression analysis.

The large positive intercept of eq 1 suggests that the substituents, with the exception of 4-OH, have relatively little effect on the activity of the model compound. The negative coefficient of I-1 proves that the 4-OH group greatly reduces the inhibition potency. The positive slope with I-2 suggests that substituents, such as CH_3 and Cl, at the 5 position will increase the activity slightly and the

Table I. I_{50} Values of Quinazolines as Inhibitors for Dihydrofolate Reductase from Human Acute Lymphocytic Leukemia (ALL) and Mouse L1210R Leukemic Cells

	substituents				indicator variables and physicochem param											log ($1/I_{50}$) for DHFR			
																ALL		L1210R	
compd	X	Y	Z	R	*I*-1	*I*-2	*I*-3	*I*-4	*I*-5	*I*-6	*I*-7	π-6	MR-6	π-R	MR-R	obsd	pred	obsd	pred
1	NH_2	H	CH_2NH	4′-$CONHCH(CO_2H)(CH_2)_2CO_2H$	0	0	0	0	0	0	0	1.00	3.47	−1.81	3.51	9.21	8.65		
2	NH_2	H	CH_2NH	4′-$CONHCH(CO_2H)CH_2CO_2H$	0	0	0	0	0	1	0	1.00	3.47	−2.31	3.05	8.94	8.88	8.76	8.59
3	NH_2	H	CH_2NH	4′-$CONHCH(COOH)_2$	0	0	0	0	0	1	0	1.00	3.47	−2.81	2.55	9.10	9.12	8.81	8.65
4	NH_2	H	CH_2NH	4′-$CONHCH(CH_2COOH)_2$	0	0	0	0	0	1	0	1.00	3.47	−1.81	3.54	8.25	8.64	8.14	8.54
5	NH_2	H	CH_2NH	4′-$CONHCH(CO_2H)(CH_2)_2CONH_2$	0	0	0	0	0	1	0	1.00	3.47	−2.85	3.67	10.05	8.92	9.66[a]	8.66[b]
6	NH_2	H	CH_2NH	2′-Cl, 4′-$CONHCH(CO_2H)CH_2CO_2H$	0	0	0	0	0	1	1	1.00	3.47	−1.60	3.65	8.82	8.56	9.13[a]	8.51[b]
7	NH_2	H	CH_2NH	2′-Cl, 5′-$CONHCH(CO_2H)CH_2CO_2H$	0	0	0	0	0	1	1	1.00	3.47	−1.60	3.65	8.52	8.56	8.54	8.51
8	NH_2	H	CH_2NH	4′-Cl, 3′-$CONHCH(CO_2H)CH_2CO_2H$	0	0	0	0	0	1	1	1.00	3.47	−1.60	3.65	7.92	8.56	8.37	8.51
9	NH_2	H	CH_2NH	4′-Cl, 3′-$CONHCH(CO_2H)(CH_2)_2CO_2H$	0	0	0	0	0	0	1	1.00	3.47	−1.10	4.11	6.70[a]	8.33[b]	8.03	8.45
10	NH_2	H	CH_2NH	3′-Br	0	0	0	0	1	1	1	1.00	3.47	0.86	0.89	8.66	8.36	8.64	8.23
11	NH_2	H	CH_2NH	4′-Cl	0	0	0	0	1	1	0	1.00	3.47	0.71	0.60	8.22	8.46	8.30	8.25
12	NH_2	H	CH_2NH	3′-Cl	0	0	0	0	1	1	1	1.00	3.47	0.71	0.60	8.00	8.46	8.00	8.25
13	NH_2	H	CH_2NH	3′,4′-Cl_2	0	0	0	0	1	1	1	1.00	3.47	1.42	1.20	8.42	8.14	8.05	8.16
14	NH_2	H	CH_2NH	3′-CF_3, 4′-Cl	0	0	0	0	1	1	1	1.00	3.47	1.59	1.10	8.40	8.11		
15	NH_2	H	CH_2NH	3′,4′,5′-$(OCH_3)_3$	0	0	0	0	1	1	1	1.00	3.47	−0.06	2.36	8.22	8.35		
16	NH_2	Cl	CH_2NH	4′-$CONHCH(CO_2H)CH_2CO_2H$	0	1	0	0	0	1	0	1.00	3.47	−2.31	3.05	9.35	9.17	9.21	9.09
17	NH_2	Cl	CH_2NH	4′-$CONHCH(CO_2H)(CH_2)_2CO_2H$	0	1	0	0	0	0	0	1.00	3.47	−1.81	3.51	9.11	8.94	9.05	9.03
18	NH_2	Cl	CH_2NH	3′,4′-Cl_2	0	1	0	0	1	1	1	1.00	3.47	1.42	1.20	8.96	8.43	8.70	8.66
19	NH_2	CH_3	CH_2NH	4′-$CONHCH(CO_2H)CH_2CO_2H$	0	1	0	0	0	1	0	1.00	3.47	−2.31	3.05	8.90	9.17	8.87	9.09
20	NH_2	CH_3	CH_2NH	4′-$CONHCH(CO_2H)(CH_2)_2CO_2H$	0	1	0	0	0	0	0	1.00	3.47	−1.81	3.51	8.78	8.94		
21	NH_2	CH_3	CH_2NH	4′-$CONHCH(CO_2Et)(CH_2)_2CO_2Et$	0	1	0	0	0	1	0	1.00	3.47	−0.81	5.36	8.49	8.29		
22	NH_2	CH_3	CH_2NH	4′-$CONHCH(CH_2CO_2H)_2$	0	1	0	0	0	1	0	1.00	3.47	−1.81	3.54	8.17	8.93	9.15	9.03
23	NH_2	CH_3	CH_2NH	4′-Cl, 3′-$CONHCH(CO_2H)(CH_2)_2CO_2H$	0	1	0	0	0	0	1	1.00	3.47	−1.10	4.11	8.52	8.62	8.68	8.95
24	NH_2	CH_3	CH_2NH	3′-Br	0	1	0	0	1	1	1	1.00	3.47	0.86	0.89	8.82	8.65	8.82	8.72
25	NH_2	CH_3	CH_2NH	3′,4′-Cl_2	0	1	0	0	1	1	1	1.00	3.47	1.42	1.20	8.59	8.43		
26	NH_2	CH_3	CH_2NH	3′,4′,5′-$(OCH_3)_3$	0	1	0	0	1	1	1	1.00	3.47	−0.06	2.36	8.89	8.64	8.92	8.83
27	NH_2	H	CH_2NCH_3	4′-$CONHCH(CO_2H)CH_2CO_2H$	0	0	0	1	0	1	0	2.09	3.96	−2.31	3.05	8.69	8.69	8.66	8.45
28	NH_2	H	CH_2NCH_3	4′-$CONHCH(CO_2H)(CH_2)_2CO_2H$	0	0	0	1	0	0	0	2.09	3.96	−1.81	3.51	8.41	8.46	8.80	8.39
29	NH_2	H	CH_2NCH_3	3′,4′-Cl_2	0	0	0	1	1	1	1	2.09	3.96	1.42	1.20	8.30	7.95		
30	NH_2	H	$CH_2N(CHO)$	4′-$CONHCH(CO_2H)(CH_2)_2CO_2H$	0	0	0	1	0	0	0	0.49	3.96	−1.81	3.51	8.22	8.46		
31	NH_2	H	$CH_2N(NO)$	3′,4′-Cl_2	0	0	0	1	1	1	1	1.25	3.99	1.42	1.20	8.00	7.94	8.00	8.00
32	NH_2	H	$CH_2N(COCH_3)$	3′,4′-Cl_2	0	0	0	1	1	1	1	0.09	4.42	1.42	1.20	7.30	7.78		
33	NH_2	H	$OC_6H_4CH_2NH$	4′-$CONHCH(CO_2H)(CH_2)_2CO_2H$	0	0	0	1	0	0	0	3.08	6.24	−1.81	3.51	7.80	7.59	7.71	7.71
34	NH_2	H	$OC_6H_4CH_2NH$	4′-$CONHCH(CO_2H)CH_2CO_2H$	0	0	0	1	0	1	0	3.08	6.24	−2.31	3.05	7.77	7.82	7.67	7.77
35	NH_2	CH_3	$CH_2N(NO)$	3′,4′-Cl_2	0	1	0	1	1	1	1	1.25	3.99	1.42	1.20	8.00	8.23		
36	NH_2	CH_3	$NHCH_2$	4′-$CONHCH(CO_2H)(CH_2)_2CO_2H$	0	1	1	0	0	0	0	1.00	3.47	−1.81	3.51	9.07	8.94		
37	NH_2	CH_3	$NHCH_2$	4′-$COOC_2H_5$	0	1	1	0	1	1	0	1.00	3.47	0.51	1.75			8.00[a]	8.46[b]
38	NH_2	H	$NHCH_2$	4′-$CONHCH(CO_2H)(CH_2)_2CO_2H$	0	0	1	0	0	0	0	1.00	3.47	−1.81	3.51	8.00	8.65		
39	NH_2	H	$NHCH_2$	4′-$CONHCH(CO_2Et)(CH_2)_2CO_2Et$	0	0	1	0	0	1	0	1.00	3.47	−0.81	5.36	7.66	8.00		
40	NH_2	H	$NHCH_2$	4′-$CONHCH(CO_2H)CH_2CO_2H$	0	0	1	0	0	1	0	1.00	3.47	−2.31	3.05			7.36	7.36
41	OH	H	CH_2CH_2	4′-$CONHCH(CO_2H)(CH_2)_2CO_2H$	1	0	0	1	0	0	0	2.66	3.47	−1.81	3.51	6.31	5.78		
42	OH	H	CH_2NCH_3	4′-$CONHCH(CO_2H)(CH_2)_2CO_2H$	1	0	0	1	0	0	0	2.09	3.96	−1.81	3.51	6.22	5.59		
43	OH	H	CH_2S	4′-$CONHCH(CO_2Et)(CH_2)_2CO_2Et$	1	0	0	1	0	1	0	2.82	3.81	−0.81	5.36	5.05	5.00		
44	OH	H	CH_2S	4′-$CONHCH(CO_2H)(CH_2)_2CO_2H$	1	0	0	1	0	0	0	2.82	3.81	−1.81	3.51	4.90	5.65		
45	OH	H	CH_2NH	4′-$COOC_2H_5$	1	0	0	0	1	1	0	1.00	3.47	0.51	1.75	4.66	5.43		
46	OH	H	$NHCH_2$	4′-$CONHCH(CO_2H)(CH_2)_2CO_2H$	1	0	1	0	0	0	0	1.00	3.47	−1.81	3.51	5.78	5.78		
47	OH	H	$NHCH_2$	4′-$CONHCH(CO_2Et)(CH_2)_2CO_2Et$	1	0	1	0	0	1	0	1.00	3.47	−0.81	5.36	5.65	5.13		
48	OH	H	$NHCH_2$	4′-$COOC_2H_5$	1	0	1	0	1	1	0	1.00	3.47	0.51	1.75	5.49	5.43		
49	OH	CH_3	$NHCH_2$	4′-$CONHCH(CO_2H)(CH_2)_2CO_2H$	1	1	1	0	0	0	0	1.00	3.47	−1.81	3.51	6.12	6.07		
50	OH	CH_3	$NHCH_2$	4′-$CONHCH(CO_2Et)(CH_2)_2CO_2Et$	1	1	1	0	0	1	0	1.00	3.47	−0.81	5.36	5.10	5.42		

[a] These data are not included in regression analysis. [b] Calculated from correlation equations.

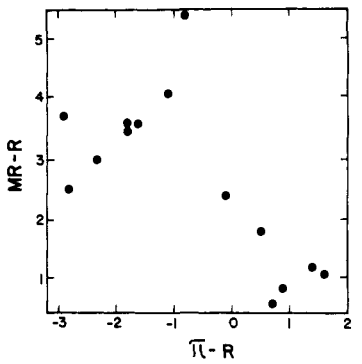


Figure 1. Illustration of the interdependence between π-R and MR-R.

magnitude of the MR-6 indicates steric sensitivity in positions 9 and 10 of substituent Z. Since π-R values are negative (except for R = $COOC_2H_5$), a smaller π-R value means a higher activity for the compound as the coefficient of this term is negative. The molar refractivity of the R group appears to have the same effect as the hydrophobicity on the inhibition potency of the substances.

By using a similar approach, the QSAR for DHFR from leukemic mice cells was also established. The molar concentration of quinazolines causing 50% inhibition of DHFR from L1210R is expressed with the substituent parameters by eq 2. Omitting compounds **5**, **6**, and **37**,

$$\begin{array}{lllll} \log(1/I_{50}) = 9.60 + 0.44(I\text{-}2) - 1.06(I\text{-}3) - & & & & \\ \quad \text{SE} & 0.36 & 0.14 & 0.24 & \\ \quad t \text{ stat} & 26.90 & 3.16 & -4.39 & \\ & & & 0.35(\text{MR-6}) - 0.145(\pi\text{-R}) & (2) \\ & & & 0.09 & 0.045 \\ & & & -3.83 & -3.21 \end{array}$$

$$n = 27;\ r = 0.83;\ s = 0.33;\ F_{4,22} = 12.51$$

which have a relatively higher deviation from calculated values, yields eq 3. Since none of the compounds in-

$$\begin{array}{lllll} \log(1/I_{50}) = 9.36 + 0.49(I\text{-}2) - 1.23(I\text{-}3) - & & & & \\ \quad \text{SE} & 0.27 & 0.11 & 0.25 & \\ \quad t \text{ stat} & 34.98 & 4.61 & -4.92 & \\ & & & 0.296(\text{MR-6}) - 0.115(\pi\text{-R}) & (3) \\ & & & 0.068 & 0.035 \\ & & & -4.38 & -3.26 \end{array}$$

$$n = 24;\ r = 0.904;\ s = 0.235;\ F_{4,19} = 21.16$$

vestigated have a 4-OH substituent, the indicator variable I-1 no longer appears in eq 2. The large negative coefficient of I-3 indicates that Z = $-NHCH_2-$ will greatly decrease inhibition potency. As discussed above, the substituent in the 5 position of quinazolines is the influential factor for the inhibition of dihydrofolate reductase from mice leukemic cells. The steric effect of the Z group (MR-6) will decrease the inhibition of DHFR from ALL and from L1210R by quinazolines. Generally, the more hydrophobic are the substituents in the phenyl ring, the lower is the inhibitory effect of the drug. QSAR formulated by Hansch et al.[17] for dihydrofolate reductase from rat liver is similar to our correlation equations; in both cases, 4-OH decreases the potency of quinazolines. On the other hand, however, we found that an increase in molar refractivity and the replacement of the "unnatural" bridge $-NHCH_2-$ in Z decrease the biological activity of quinazolines for the types of enzyme investigated. Nevertheless, we can draw the general conclusion from the two studies that quinazolines have a similar pattern of inhibition potency for mammalian DHFR from different sources.

Quinazolines as Inhibitors of Thymidylate Synthetase

The antitumor activity of the quinazolines is expected to manifest itself in their inhibition potency for thymidylate synthetase. Consequently, the corresponding QSAR is of considerable interest. Data obtained by Scanlon et al.[5] for the inhibition of thymidylate synthetase from mice L1210S leukemic cells by quinazolines were first subjected to multiple regression analysis as described before for the inhibition of DHFR. The equation had high correlation with biological activity; however, it did not offer a physical interpretation of the coefficients. This is due to the fact that multiple linear-regression analysis, the basis of the Hansch approach to QSAR, assumes either independence and noncollinearity of the predictor variables or additivity of the substituents's contributions. If one or more of these assumptions are not satisified in practice, the correlation equation may be devoid of physical meaning and a trial and error approach is required to formulate a meaningful correlation.

On the other hand, by using multivariate statistics such as cluster analysis, factor analysis, and discriminant analysis, the interdependence of the variables can be conveniently tested and thereby the formulation of an appropriate correlation equation facilitated. Multivariate analysis sheds light on the structure and elucidates the main features of a data matrix. It can involve different statistical procedures; each of them furnishes somewhat different types of information about the data. A detailed description of multivariate analysis can be found in most statistical texts.[20,21] Computer programs for performing multivariate analysis are generally available, e.g., SPSS,[22] BMDP,[23] and SAS.[24]

In this study we applied this technique to establish QSAR for the quinazolines listed in Table II with regard to the inhibition of thymidylate synthetase.[5] Before presenting the results, however, we shall briefly examine and illustrate the salient features of the individual statistical approaches in the light of the inhibition of thymidylate synthetase by quinazolines.

Multivariate Analysis of the Relationships between Physicochemical Parameters and the Drug Potency. (a) Cluster Analysis. Cluster analysis[25] is one of the pattern-recognition techniques[26] and searches for similarities of objects in the data matrix. First, a "distance matrix" is generated and then the closest pairs of objects are combined into clusters. The objects of clustering can be either cases (compounds) or variables (physicochemical parameters or indicator variables).

Cluster analysis of variables provides criteria for the selection of the appropriate variables for multiple regression analysis when there is an appreciable redundancy of variables. The computer program BMDP1M[23] was used for clustering the variables to measure their similarity by the magnitude of the correlation coefficients listed in Table III. The tree diagram shown in Figure 2 illustrates how the clusters are formed at each step by using an amalgamation rule based on the maximum similarity over all pairings of the variables between the two clusters. The diagram clearly shows the relationships between the indicator variables and physicochemical parameters used in the study.

Cluster analysis of congeners has been investigated by Hansch et al.[13] and they found that clustering facilitates a rational selection of substituents for drugs to be synthetized, especially when the pertinent QSAR is not

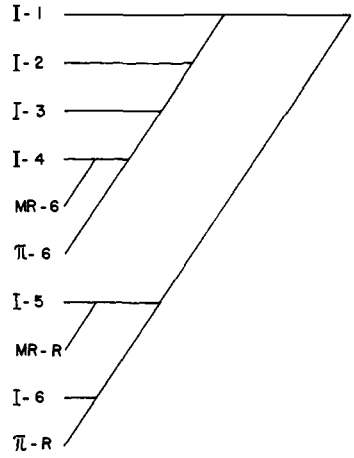


Figure 2. Tree diagram illustrating the relationship between substituent parameters as a result of cluster analysis.

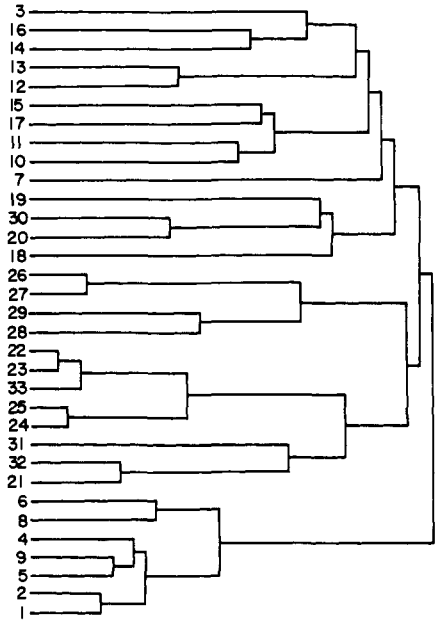


Figure 3. Hierarchical representation of the results of cluster analysis of substances.

known. Once the substances have been clustered on the basis of the physicochemical parameters, with sufficient experience in bioorganic reaction mechanism, drug metabolism, and organic synthesis, one can decide which derivatives should next be prepared.[13]

In our cluster analysis of cases, the congeners of quinazolines were clustered by using the Euclidean distance;[20] computer program BMDP2M[23] was employed. Variables which were found to be not independent in clustering were eliminated. The results are shown in Figure 3 in the form of a hierarchial diagram. There are three major clusters: compounds with an amide function attached to the ring, substances with a carboxylic or ester group in the benzene ring, and congeners containing amide functional groups but having a Z group other than $-CH_2NH-$ or $-NHCH_2-$. Consequently, the substances investigated can be divided according to their chemical nature into three groups having different potency for inhibition. Only three compounds (**20**, **30**, and **32**) are inconsistent with the observed activity; two of them (**20** and **32**) may be due to experimental error according to the correlation eq 7, as shown later. The cutoff point for inhibition potency is $I_{50} = 10^{-5}$ M.

Cluster analysis enables us to classify closely related congeners into distinct categories according to their potency and substituents and to establish a certain structure–activity relationship. Often cluster analysis alone does not provide enough quantitative information for QSAR, and other multivariate statistical methods such as factor analysis should also be employed.

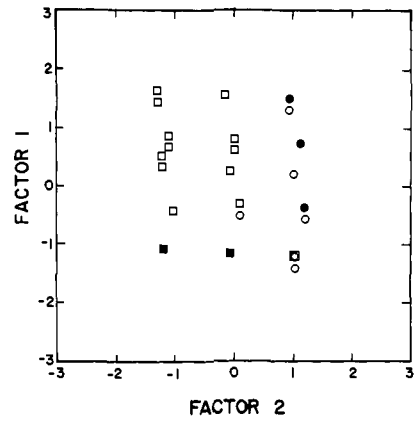


Figure 4. Illustration of the factor scores of all quinazolines for factors 1 and 2. The open symbols □ and ○ represent high and low potent compounds, respectively. Overlap is indicated by solid symbols.

(b) Factor Analysis. Originally, factor analysis was developed in the field of educational psychology.[27] It is used to explain a set of multivariate data in terms of a fairly small number of underlying factors and to analyze the dependence of each factor on the variables.[28] Factor analysis has been used previously in structure–activity studies[11] and as a preprocessing method to separate compounds into classes on the basis of their molecular descriptor coding,[12] which correspond to our indicator variables.

There are four main steps in factor analysis. First, all variables are correlated by clustering them into factors in such a way that the pertinent variables are highly correlated. Second, the process of initial factor extraction is carried out by estimating the factor loadings, λ_{ij}, using eq

$$z_i = \sum_{j=1}^{m} \lambda_{ij} f_j + e_i \qquad (4)$$

4 where f_j denotes the common factors and e_i is the factor unique to variable z_i. Third, the factors are rotated to make the loading for each factor either large or small, and thereby they facilitate the physical interpretation of the factors. Fourth, the factor scores are computed from eq

$$f_j = \sum_{i=1}^{p} b_{ji} z_i \qquad (5)$$

5 where b_{ji} are the factor score coefficients.

Factor analysis was performed by using the indicator variables listed in Table II. The correlation matrix obtained in the first step is shown in Table III. For initial factor extraction, we selected principal component analysis,[28] whereas in the third step of factor analysis the varimax method of orthogonal rotation[28] was chosen. As a result, three factors were extracted with a cumulative proportion of total variance of 0.766 and the eigenvalue of the third factor was 0.995. Table IV lists the rotated factor loadings, and the pertinent factor score coefficients are given in Table V. The tabulated results demonstrate that the most important variables in factor 1 are *I*-2, *I*-3, and *I*-4. Factor 2 essentially depends on *I*-5 and *I*-6, and factor 3 is nearly identical to *I*-1.

The plot of factor scores in Figure 4 gives us a graphical representation of some of the results by showing the factor scores of all compounds for factors 1 and 2. Similar graphs

Table II. Inhibition Potency and Physicochemical Parameters for Inhibition of Thymidylate Synthetase by Quinazolines[d]

	structure				indicator variables and physicochemical parameters											log $(1/I_{50})$			
																L1210S		*L. casei*	
compd	X	Y	Z	R'[a]	*I*-1	*I*-2	*I*-3	*I*-4	*I*-5	*I*-6	π-5	MR-5	π-6	MR-6	π-R	MR-R	obsd	pred	obsd	pred
1	OH	H	CH_2NCH_3	Glu	1	0	0	1	0	0	0.00	0.10	2.09	3.96	−1.81	3.51	7.30	6.74	6.40[b]	5.13[c]
2	OH	H	$CH_2N(CHO)$	Glu	1	0	0	1	0	0	0.00	0.10	0.49	3.96	−1.81	3.51	7.00	6.74	5.52[b]	4.45[c]
3	OH	CH_3	CH_2NH	Glu	1	1	0	0	0	0	0.56	0.57	1.00	3.47	−1.81	3.51	7.00	5.98	4.90	4.67
4	OH	H	CH_2S	Glu	1	0	0	1	0	0	0.00	0.10	2.82	3.81	−1.81	3.51	6.00	6.74	5.60	5.44
5	OH	H	CH_2CH_2	Glu	1	0	0	1	0	0	0.00	0.10	2.66	3.47	−1.81	3.51	5.30[b]	6.74[c]	5.12	5.38
6	OH	H	CH_2S	$Glu(Et)_2$	1	0	0	1	0	1	0.00	0.10	2.82	3.81	−0.81	5.36	7.00	6.74	5.10	4.63
7	OH	H	CH_2NH	$Glu(Et)_2$	1	0	0	0	0	1	0.00	0.10	1.00	3.47	−0.81	5.36	6.60	5.98	4.15	3.86
8	OH	H	CH_2NCH_3	$Glu(Et)_2$	1	0	0	1	0	1	0.00	0.10	2.09	3.96	−0.81	5.36	5.12[b]	6.74[c]	4.12	4.32
9	OH	H	CH_2O	Glu	1	0	0	1	0	0	0.00	0.10	1.66	3.22	−1.81	3.51	5.00[b]	6.74[c]	4.00[b]	4.95[c]
10	OH	H	$NHCH_2$	Glu	1	0	1	0	0	0	0.00	0.10	1.00	3.47	−1.81	3.51	6.30	5.98	5.00	5.06
11	OH	H	$NHCH_2$	$Glu(Et)_2$	1	0	1	0	0	1	0.00	0.10	1.00	3.47	−0.81	5.36	6.00	5.98	4.46	4.25
12	OH	CH_3	$NHCH_2$	Glu	1	1	1	0	0	0	0.56	0.57	1.00	3.47	−1.81	3.51	5.12	5.98	4.30[b]	5.06[c]
13	OH	CH_3	$NHCH_2$	$Glu(Et)_2$	1	1	1	0	0	1	0.56	0.57	1.00	3.47	−0.81	5.36	5.12	5.98	4.00	4.25
14	NH_2	CH_3	$NHCH_2$	Glu	0	1	1	0	0	0	0.56	0.57	1.00	3.47	−1.81	3.51	6.60	5.98	5.60	5.45
15	NH_2	H	$NHCH_2$	Glu	0	0	1	0	0	0	0.00	0.10	1.00	3.47	−1.81	3.51	6.60	5.98	5.30	5.45
16	NH_2	CH_3	CH_2NH	Glu	0	1	0	0	0	0	0.56	0.57	1.00	3.47	−1.81	3.51	6.00	5.98	5.00	5.06
17	NH_2	H	$NHCH_2$	$Glu(Et)_2$	0	0	1	0	0	1	0.00	0.10	1.00	3.47	−0.81	5.36	6.00	5.98	5.00	4.65
18	NH_2	CH_3	CH_2NH	$Glu(Et)_2$	0	1	0	0	0	1	0.56	0.57	1.00	3.47	−0.81	5.36	5.12	5.98	4.30	4.25
19	NH_2	H	CH_2NH	Glu	0	0	0	0	0	0	0.00	0.10	1.00	3.47	−1.81	3.51	5.00	5.98	4.00[b]	5.06[c]
20	NH_2	H	CH_2NH	$Glu(Et)_2$	0	0	0	0	0	1	0.00	0.10	1.00	3.47	−0.81	5.36	4.00[b]	5.98[c]	4.00	4.25
21	OH	H	CH_2NCH_3	OH	1	0	0	1	1	1	0.00	0.10	2.09	3.96	−0.32	0.69	4.70	4.72	4.12	4.32
22	OH	H	CH_2NH	OH	1	0	0	0	1	1	0.00	0.10	1.00	3.47	−0.32	0.69	4.00	3.97	4.00	3.86
23	OH	H	CH_2NH	OC_2H_5	1	0	0	0	1	1	0.00	0.10	1.00	3.47	0.51	1.75	4.00	3.97	4.00	3.86
24	OH	H	$NHCH_2$	OH	1	0	1	0	1	1	0.00	0.10	1.00	3.47	−0.32	0.69	4.00	3.97	4.00	4.25
25	OH	H	$NHCH_2$	OC_2H_5	1	0	1	0	1	1	0.00	0.10	1.00	3.47	0.51	1.75	4.00	3.97	4.30	4.25
26	OH	CH_3	$NHCH_2$	OH	1	1	1	0	1	1	0.56	0.57	1.00	3.47	−0.32	0.69	4.05	3.97	4.05	4.25
27	OH	CH_3	$NHCH_2$	OC_2H_5	1	1	1	0	1	1	0.56	0.57	1.00	3.47	0.51	1.75	4.00	3.97	4.05	4.25
28	NH_2	CH_3	$NHCH_2$	OC_2H_5	0	1	1	0	1	1	0.56	0.57	1.00	3.47	0.51	1.75	4.00	3.97	5.00	4.65
29	NH_2	CH_3	CH_2NH	OC_2H_5	0	1	0	0	1	1	0.56	0.57	1.00	3.47	0.51	1.75	4.00	3.97	4.00	4.25
30	NH_2	H	CH_2NH	OC_2H_5	0	0	0	0	1	1	0.00	0.10	1.00	3.47	0.51	1.75	4.00	3.97	4.00	4.25
31	NH_2	H	$CH_2N(CHO)$	OH	0	0	0	1	1	1	0.00	0.10	0.49	3.96	−0.32	0.69	4.00	4.72	4.12	4.04
32	SH	H	CH_2NCH_3	OH	1	0	0	1	1	1	0.00	0.10	2.09	3.96	−0.32	0.69	5.12	4.72	4.12	4.32
33	SH	H	CH_2NH	OH	1	0	0	0	1	1	0.00	0.10	1.00	3.47	−0.32	0.69	4.00	3.97	4.00	3.86

[a] R group was represented by $-C(=O)R'$ and attached to the para position of the phenyl ring. [b] These data are not included in regression analysis. [c] Calculated from correlation equations. [d] The enzyme was obtained from mouse leukemic cell line L1210S and from *Lactobacillus casei*.

Table III. Correlation Matrix

	I-1	I-2	I-3	I-4	I-5	I-6	π-6	MR-6	π-R	MR-R
I-1	1.000									
I-2	−0.233	1.000								
I-3	0.000	0.324	1.000							
I-4	0.326	−0.435	−0.498	1.000						
I-5	0.044	0.008	0.035	−0.127	1.000					
I-6	0.000	−0.050	0.048	−0.187	0.609	1.000				
π-6	0.378	−0.302	−0.347	0.695	−0.201	−0.137	1.000			
MR-6	0.198	−0.328	−0.376	0.755	0.043	0.048	0.423	1.000		
π-R	−0.050	0.050	0.087	−0.257	0.842	0.867	−0.214	−0.049	1.000	
MR-R	−0.100	0.012	0.010	−0.003	−0.887	−0.210	0.147	−0.090	−0.501	1.000

Table IV. Rotated Factor Loadings

	factor		
parameter	1	2	3
I-1	−0.125	0.019	0.941
I-2	0.684	−0.089	−0.290
I-3	0.852	0.015	0.263
I-4	−0.780	−0.194	0.309
I-5	0.028	0.887	0.035
I-6	0.031	0.900	−0.017

Table V. Factor Score Coefficients

	factor		
parameter	1	2	3
I-1	0.087	0.001	0.854
I-2	0.355	−0.087	−0.151
I-3	0.543	−0.043	0.391
I-4	−0.393	−0.081	0.157
I-5	−0.026	0.543	0.022
I-6	−0.034	0.551	−0.027

can be plotted for other pairs of factors as well. Our data points represent either "high potent" or "low potent" substances, and the results show that compounds having subzero values for factor 2, which is essentially a linear combination of I-5 and I-6, are the strongest inhibitors for thymidylate synthetase from mice L1210S leukemia. Only two observed high potent compounds, **7** and **32**, have positive scores for factor 2: 0.122 and 1.041, respectively. The latter has been predicted as a low potent compound by correlation eq 7. Other similar plots, however, indicate that factors 2 and 3 are not critical with respect to the potency of the drugs according to this analysis. The results exemplify that factor analysis can be of use in establishing structure–activity relationships, although it remains an essentially preprocessing method before a complete QSAR is established because the abstract factors do not express an explicit relationship between biological activity and physicochemical parameters. In the next section, we illustrate the use of discriminant analysis, which yields further information and facilitates the establishment of QSAR.

(c) Discriminant Analysis. This multivariate analytical method investigates the relationship between a known grouping of the data and the variables by generating a classification function that maximizes group differences. It is used to assign individual compounds to those groups to which they should belong according to their chemical structure. In our analysis the BMDP7M[22] stepwise discriminant analysis program was used. 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 classification scheme. Should the F value for a particular variable become too small as other variables are added that variable is eliminated from the classification.

The computation procedures are documented in BMDP manuals. The Mahalanobis distances D^2 (the distance from each case to each group mean)[20] are computed for each compound, outliers can be identified as cases with large D^2 from their group means. For the quinazolines under investigation, it has been found that the two structural variables I-4 and I-5 suffice to yield a classification function to assign high or low potency to the compounds. The classification function for low potency compounds ($I_{50} > 10^{-5}$ M) is given by

$$-1.09(I\text{-}4) + 15.72(I\text{-}5) - 8.10$$

and for high potency compounds ($I_{50} \leq 10^{-5}$ M) is expressed as

$$1.90(I\text{-}4) + 0.06(I\text{-}5) - 0.88$$

The classification function places 31 out of 33 compounds into the correct category at the 1% significance level ($F_{2,30} = 54.9$). These two outliers (compounds **20** and **32**) probably are caused by erroneous experimental results, which will be discussed in the section on regression analysis of correlation equations.

Discriminant analysis of the data in the L1210S system indicates that the carboxylic or ester group in the benzene ring, as well as the Z group in position 6, has the greatest influence on the inhibitory effect of the quinazolines investigated.

Combination of Multivariate Statistical Methods. It has been shown above that cluster analysis, factor analysis, and discriminant analysis can individually offer another approach to QSAR. According to discriminant analysis, the two indicator variables I-4 and I-5 enable us to classify the quinazoline congeners with 94% accuracy into two distinct categories: high potent and low potent compounds. Factor analysis showed that substances with negative values of factor 2, which contains I-5 and I-6, are highly potent inhibitors for thymidylate synthetase from the L1210S system. The compounds investigated can be clustered into three categories, as seen in cluster analysis.

In the light of the above discussion, we suggest that the various multivariate statistical methods be used for the establishment of QSAR according to the scheme depicted in Figure 5. At first, cluster analysis can be employed to find the interdependence of the variables and thereby to eliminate redundant variables. Whereas factor analysis can provide a primary structure–activity relationship, this technique does not necessarily yield information directly useful for setting up the correlation equation. On the other hand, discriminant analysis can be invaluable in affording a pattern for the correlation equation. In fact, eq 6 to 9 are directly patterned after the results of discriminant analysis.

Multiple Regression and Correlation Equations. By choosing the variables according to the information obtained from multivariate analyses and implementing the leaps and bounds algorithm for best subset regression, the

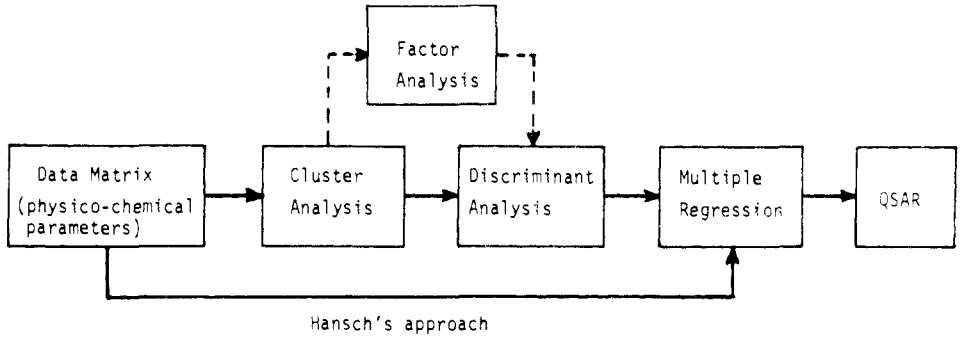


Figure 5. Path proposed for the use of various multivariate statistical methods in the establishment of QSAR.

correlation equation for the inhibition of thymidylate synthetase from mice L1210S leukemic cells was obtained by eq 6.

$$\log (1/I_{50}) = 5.77 + 0.40(I\text{-}4) - 1.72(I\text{-}5) \quad (6)$$

SE	0.19	0.27	0.26
t-stat	30.65	1.45	-6.62

$n = 33$; $r = 0.788$; $s = 0.72$; $F_{2,30} = 24.6$

The correlation between biological activity and chemical structure in the above "complete" equation was poor. Deleting compounds **5**, **8**, **9**, and **20** in the regression analysis, we find a better correlation equation given by eq 7.

$$\log (1/I_{50}) = 5.98 + 0.75(I\text{-}4) - 2.01(I\text{-}5) \quad (7)$$

SE	0.14	0.23	0.20
t stat	41.80	3.30	-10.27

$n = 29$; $r = 0.905$; $s = 0.525$; $F_{2,26} = 58.96$

The predicted inhibition potency based on correlation eq 7 is shown in Table II. Compound **32** has the low potency and is consistent with the results obtained by multivariate analysis. The large coefficient of the term I-5 indicates that the R group in the benzene ring is the decisive factor in determining inhibition activity. The Z group in position 6 of quinazolines is the next important parameter which influences the potency, as suggested by the coefficient of I-4.

In a similar fashion, QSAR for the inhibition of thymidylate synthetase from *Lactobacillus casei* can be formulated by using multivariate statistical methods and multiple regression analysis. The correlation equation is given by eq 8.

$$\log (1/I_{50}) = 4.59 - 0.25(I\text{-}1) + 0.29(I\text{-}3) - 0.77(I\text{-}6) + 0.39(\pi\text{-}6) \quad (8)$$

SE	0.27	0.20	0.19	0.18	0.16
t stat	16.75	-1.29	1.56	-4.38	2.41

$n = 33$; $r = 0.71$; $s = 0.48$; $F_{4,28} = 7.21$

Upon omission of compounds **1**, **2**, **9**, **12**, and **19** in the regression analysis, QSAR can be significantly improved as seen from eq 9.

$$\log (1/I_{50}) = 4.638 - 0.395(I\text{-}1) + 0.391(I\text{-}3) - 0.809(I\text{-}6) + 0.426(\pi\text{-}6) \quad (9)$$

SE	0.173	0.110	0.105	0.113	0.093
t stat	26.74	-3.60	3.72	-7.19	4.58

$n = 28$; $r = 0.911$; $s = 0.246$; $F_{4,23} = 28$

The results show that compounds with 4-NH_2 and glutamic acid in substituent R are highly potent inhibitors because the coefficients of I-1 and I-6 are negative. Quinazolines with Z = $-NHCH_2-$ have a little higher inhibitory activity than those with Z = $-CH_2NH-$ for thymidylate synthetase from *Lactobacillus casei*, since the coefficient of I-3 in eq 9 is positive. The hydrophobic properties of the substituent Z (π-6), also play an important role in determining the inhibitory properties.

Conclusion

QSAR for quinazolines as the inhibitors of dihydrofolate reductase and thymidylate synthetase from different sources have been studied. Multivariate statistics including discriminant analysis, factor analysis, and cluster analysis have been found to facilitate the development of a satisfactory correlation equation. The inhibition of bacterial thymidylate synthetase by quinazolines yields QSAR significantly different from that for the same enzyme from mammalian sources. This finding suggests that there are significant differences between the two types of the enzyme.

Acknowledgment. We thank Jean-Fern Chen for her valuable assistance in the application and interpretation of multivariate statistical methods. This study was supported by Grant CA 21948 from the National Cancer Institute, U.S. Public Health Service, DHEW.

References and Notes

(1) J. R. Bertino, *Handb. Exp. Pharmakol.*, **38**, 468–483 (1975).
(2) B. R. Barker and B. T. Ho, *J. Pharm. Sci.*, **53**, 1137 (1964).
(3) (a) C. Hansch, C. Silipo, and E. E. Steller, *J. Pharm. Sci.*, **64**, 1186 (1975); (b) C. Silipo and C. Hansch, *J. Am. Chem. Soc.*, **97**, 6849 (1975); (c) C. Silipo and C. Hansch, *J. Med. Chem.*, **19**, 62 (1976); (d) M. Yoshimoto and C. Hansch, *ibid.*, **19**, 71 (1976).
(4) P. V. Danenberg, *Biochim. Biophys. Acta*, **473**, 73 (1977).
(5) K. J. Scanlon, J. B. Hynes, B. A. Moroson, and J. R. Bertino, *Mol. Pharmacol.*, in press.
(6) K. J. Scanlon, W. Rode, J. B. Hynes, and J. R. Bertino, *Proc. Am. Assoc. Cancer Res.*, **19**, 541 (1978).
(7) (a) C. Hansch, *Acc. Chem. Res.*, **2**, 232 (1969); (b) C. Hansch in "Drug Design", Vol. I, E. J. Ariens, Ed., Academic Press, New York, 1971.
(8) S. M. Free and J. W. Wilson, *J. Med. Chem.*, **7**, 395 (1964).
(9) Y. C. Martin, J. B. Holland, C. H. Jarboe, and N. Plotnikoff *J. Med. Chem.*, **17**, 409 (1974).
(10) (a) E. M. Hodnett, G. Prakash, and J. Amirmoazzami, *J. Med. Chem.*, **21**, 11 (1978); (b) G. Prakash and E. M. Hodnett, *ibid.*, **21**, 369 (1978).
(11) M. L. Weiner and P. H. Weiner, *J. Med. Chem.*, **16**, 655 (1973).
(12) A. Cammarata and G. K. Menon, *J. Med. Chem.*, **19**, 739 (1976).
(13) C. Hansch, S. H. Unger, and A. B. Forsythe, *J. Med. Chem.*, **16**, 1217 (1973).
(14) (a) A. Leo, C. Hansch, and D. Elkins, *Chem. Rev.*, **71**, 525 (1971). (b) The π values presented here are calculated for neutral substituents and are relatively fair estimates because of their minor importance in this QSAR. The partition coefficients will be changed upon ionization; therefore, for

other studies, more reliable π values should be obtained preferably from experiment.
(15) (a) C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nikaitani, and E. L. Lien, *J. Med. Chem.*, **16**, 1207 (1973); (b) C. Hansch, S. D. Rockwell, P. Y. C. Jow, A. Leo, and E. E. Steller, *ibid.*, **20**, 304 (1977); (c) C. Hansch and A. Leo, Pomona College Medicinal Chemistry Project, Claremont, Calif., 1978.
(16) (a) W. T. Ashton, F. C. Walker III, and J. B. Hynes, *J. Med. Chem.*, **16**, 694 (1973); (b) J. B. Hynes, W. T. Ashton, D. Bryansmith, and J. H. Freisheim, *ibid.*, **17**, 1023 (1974); (c) J. B. Hynes, J. M. Buck, L. D'Souza, and J. H. Freisheim, *ibid.*, **18**, 1191 (1975).
(17) (a) J. Y. Fukunaga, C. Hansch, and E. E. Steller, *J. Med. Chem.*, **19**, 605 (1976); (b) C. Hansch, J. Y. Fukunaga, P. Y. C. Jow, and J. B. Hynes, *ibid.*, **20**, 96 (1977).
(18) J. R. Bertino et al., in preparation.
(19) G. M. Furnival and R. W. Wilson, *Technometrics*, **16**, 499 (1974).
(20) D. F. Morrison, "Multivariate Statistical Methods", 2nd ed, McGraw-Hill, New York, 1976.
(21) J. E. Overall and C. J. Klett, "Applied Multivariate Analysis", McGraw-Hill, New York, 1972.
(22) N. H. Nie, C. H. Hull, J. G. Jenkins, K. Steinbrenner, and D. H.. Bent, "Statistical Package for the Social Sciences", McGraw-Hill, New York, 1975.
(23) W. J. Dixon, and M. B. Brown, Ed., "Biomedical Computer Programs, P-Series", University of California Press, Berkeley, Calif., 1977.
(24) A. J. Barr, J. H. Goodnight, J. P. Sall, and J. T. Helwig, "A User's Guide to Statistical Analysis System", SAS Inc., 1976.
(25) J. A. Hartigan, "Clustering Algorithms", Wiley, New York, 1975.
(26) B. R. Kowalski and C. F. Bender, *J. Am. Chem. Soc.*, **96**, 916 (1974).
(27) C. Spearman, *Am. J. Psychol.*, **15**, 201 (1904).
(28) H. H. Harman, "Modern Factor Analysis", 3rd ed, The University of Chicago Press, Chicago, Ill., 1976.

Quantitative Structure–Activity Relationships in 2,5-Bis(1-aziridinyl)-*p*-benzoquinone Derivatives against Leukemia L-1210

Masafumi Yoshimoto,* Hachio Miyazawa, Hideo Nakao, Kenkichi Shinkai, and Masao Arakawa

Central Research Laboratories, Sankyo Company Ltd., 1-2-58, Hiromachi, Shinagawa-ku, Tokyo, 140, Japan.
Received October 20, 1978

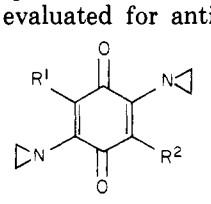
Antileukemic activities of more than 30 2,5-bis(1-aziridinyl)-*p*-benzoquinones (4) were correlated against well-defined physicochemical constants. These compounds were evaluated against lymphoid leukemia L-1210 in BDF_1 mice. The best equations obtained exhibited a linear dependence on the hydrophobic constant, π. Characteristic aspects of the equations are that the larger the relative hydrophilicity of the drugs the stronger the antileukemic activity will be and that the more hydrophilic compounds have a greater chemotherapeutic index. Steric and electronic effects were also determined to be important. Based on the correlations, three compounds (11, 15 and 19) were designed, synthesized, and biologically evaluated.

Considerable progress has been shown by Hansch and others in quantitative structure–activity relationships (QSAR) in the area of antitumor agents.[1-4] Although a number of 2,5-bis(1-aziridinyl)-*p*-benzoquionones have been synthesized and evaluated against various tumors,[5-8] no QSAR has been established.

Our laboratory has been engaged for a number of years in the search for better drugs in this series. Syntheses of *p*-benzoquinone derivatives were proceeded[5-9] in modification of mitomycin C (1), and finally the more potent and less toxic agent carboquone, 2,5-bis(1-aziridinyl)-3-[2-(carbamoyloxy)-1-methoxyethyl]-6-methyl-*p*-benzoquinone (**2**), was chosen for the market under the trade name "Esquinon"[10] (Figure 1).

In the present study, antileukemic activities of 39 2,5-bis(1-aziridinyl)-*p*-benzoquinones were correlated against well-defined physicochemical constants.

Method. The prepared 2,5-bis(1-aziridinyl)-*p*-benzoquinones (4) were evaluated for antileukemic activity



4

against lymphoid leukemia L-1210 in BDF_1 mice according to the method of CCNSC.[11,12] L-1210 cells (10^5) were intraperitoneally inoculated. The compounds were dissolved in dimethyl sulfoxide, and sterile physiological saline was added to make injectable solutions. Four kinds of biological data were obtained: minimum effective dose (MED) and optimal dose (OD) on a chronic treatment schedule (12 days, QD 1–12) and those in single injection (Q1D, day 1 only), where MED is the dose giving a 40% increase in life span (ILS) compared to the controls and OD the dose giving maximum ILS. If a dose exceeding the OD is administered, the ILS decreases. Therefore, OD/MED might be defined as a kind of chemotherapeutic index (CI). *C* in the correlation equations is the mol/kg description of MED and OD. Most of the biological data has been previously published.[5]

The substituent constants used in this work are from the compilation by Pomona College[13] or were calculated from these values. Hydrophobic constants π_1 for R^1 and π_2 for R^2 were employed. Many examples of the calculation of π values have been reported.[13b,14]

Based on the molecular refractivity constants[13a] (MR_1 for R^1 and MR_2 for R^2), R^1 and R^2 were assigned to the groups which satisfy the condition $MR_1 \leq MR_2$. We have scaled MR values by 0.1 to obtain equiscalar with π.

In this paper, terms in MR values are defined as those accounting for steric effects and not dispersion forces (in fact, molecular refractivity constants were proportional to molecular volume constants in the Lorenz–Lorenz equation[13a]) as a working hypothesis. Then, $\pi_{1,2}$ ($\pi_1 + \pi_2$) and $MR_{1,2}$ ($MR_1 + MR_2$) were used to estimate the total hydrophobicity and steric effects of R_1 and R_2.

Results and Discussion

Most of the effective compounds were substituted by alkyl functions at R^1 and R^2 (4). Since a compound with an acetyl group did not show potent activity, other de-

0022-2623/79/1822-0491$01.00/0