focus on the comparison of chemically similar analogues, where it is clear that a substantial subset of the atoms of one drug molecule match corresponding subsets in the other molecule. Distance geometry, of course, can treat such a situation, but it can also deal with chemically very dissimilar drugs. (ii) There is usually a tacit assumption that all the analogues bind in the same orientation at the receptor site, such that their corresponding atoms always occupy the same positions. In reality, however, drug molecules bind in whatever orientation and internal conformation will minimize the free energy of the drug–receptor–solvent system. The distance geometry calculation directly simulates this search for the most favorable binding mode, and rather similar compounds may bind quite differently. (iii) Other structural methods choose a particular "active conformation" for each analogue and base their relevant geometric and steric parameters on it alone. Our approach more realistically permits a flexible drug molecule to adopt whichever energetically reasonable conformation gives the best calculated binding, given the proposed site. (iv) Most methods presume that differences in binding are (to paraphrase Hopfinger[24] in his discussion of his molecular shape analysis) a smoothly varying function of differences between analogues and, indeed, should be a linear combination of molecular differences. Granted, this is often the case; however distance geometry can also model instances where a small alteraton in chemical structure gives rise to a large difference in activity. (v) Ordinarily, the drug molecules are the focus of attention, and the receptor site is described only secondarily in terms of the environment of bound ligands. Our approach, instead, devotes primary attention to building a tangible model of the site in terms of Cartesian coordinates of the site points and empirically determined contributions to the free energy of binding from the interaction between groups on ligands and site points.

Clearly, each QSAR method has at least some sets of binding data for which it works well. We claim the distance geometry approach will account for the observations on any sort of binding study, although perhaps requiring more computational effort and adjustable parameters than other methods. In addition, we claim our method will give good results on more difficult data sets, where drugs are structurally diverse, where critical steric effects cause large differences in binding for small structural changes, and where different binding modes are implicated.

Acknowledgment. This work was supported by National Institute of Arthritis, Metabolism, and Digestive Diseases, NIH, Grant 1-R01-AM28140-01.

Structure–Antitumor Activity Relationships of 9-Anilinoacridines Using Pattern Recognition

Douglas R. Henry,† Peter C. Jurs,*,† and William A. Denny‡

Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802, and Cancer Chemotherapy Research Laboratory, University of Auckland, School of Medicine, Auckland, New Zealand.
Received September 28, 1981

A pattern-recognition analysis using the ADAPT system was performed on a set of 9-anilinoacridine antitumor agents, to determine whether computer-generated descriptors could be used to separate active from inactive compounds. A training set of 213 compounds was chosen by random computer selection from a list of 776 structures. Maximal increase in life span at the LD_{10} dosage, a response which is difficult to model using traditional Hansch analysis, was used as the measure of biological activity. A set of 18 molecular descriptors, including fragment, substructure environment, and physicochemical property descriptors (molar refraction, partial electronic charge) was identified which could correctly classify 94% of the compounds in the training set (97% of active and 85% of inactive compounds). Eight of the inactive compounds that were misclassified contained amino substituents, suggesting a role for ionization. The weight vector that was obtained from the training set was applied to a prediction set of 50 compounds that were not included in the original analysis and to a set of 69 structures drawn from the recent literature. The prediction set results, ranging from 73 to 86% correct, were lower than those of the training set, but they clearly indicate that pattern-recognition techniques can be useful in the screening of proposed or already existing agents and especially useful for the identification of active compounds.

Since the turn of the century, derivatives of acridine have been used as therapeutic agents, primarily for the control of malaria (quinacrine) and bacterial infections (proflavine and acriflavine). It has been established that the primary binding site for these compounds in vivo is DNA, by the intercalation mode. It is widely and reasonably assumed that the observed biological effects result from this tight binding, although the detailed mechanism of action remains unknown.[1,2] A similar attachment, leading to the insertion or deletion of bases, has been proposed to explain the mutagenic and carcinogenic potential of acridine compounds.[3] The mutagenic activity may result from stabilization of imperfect pairing caused by single-strand slippages of the DNA.[4,5]

The antitumor potential of acridine derivatives has also been recognized for some time. This activity may result from the fragmentation of DNA, which has recently been shown to occur in tumor cells.[6] Despite much research in this area, from the earliest studies of antitumor, acridines[7,8] until the late 1960's, no definite structure–activity

† The Pennsylvania State University.
‡ University of Auckland.

(1) Adrien Albert, "The Acridines", Edwin Arnold, Ltd., London, 1966.
(2) D J. Patel, in "Drug–DNA Interactions in Solution: Acridine Mutagen, Anthracycline Antitumor, and Peptide Antibiotic Complexes", American Chemical Society, Washington, DC, 1980; *ACS Symp. Ser.*, **no. 142**, 219 (1980).
(3) Albert Lehninger, "Biochemistry", 2nd ed., Worth, New York, 1975, p 880.
(4) John R. Roth, *Annu. Rev. Genet.*, 8, 319 (1974).
(5) J. P. Schreiber and M. P. Daune, *J. Mol. Biol.*, 83, 487 (1974).
(6) N. B. Furlong, J. Sato, and T. Brown, *Cancer Res.*, 38, 1329 (1978).

0022-2623/82/1825-0899$01.25/0

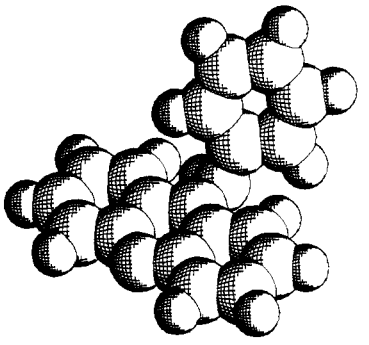
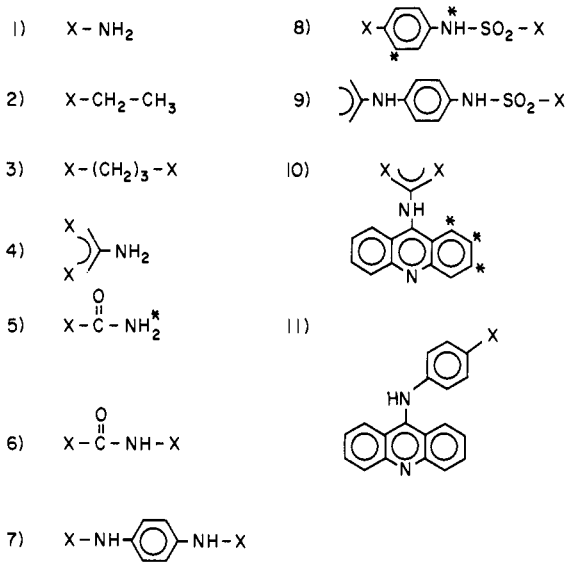


Figure 1. Computer-generated structure of 9-anilinoacridine. Binding to DNA is presumed to involve intercalation of the acridine moiety between adjacent base pairs, with the aniline ring lying in the minor groove of the DNA helix.

relationships appeared, and no antitumor agents of clinical significance were developed. In 1971, Cain and co-workers embarked on the study of some 9-anilinoacridines as antitumor agents against intraperitoneally implanted L1210 leukemia tumors in mice.[9] Based on structural requirements for a number of bis(imidazole) and quaternary ammonium heterocycles,[10] these workers have since synthesized and tested hundreds of acridine derivatives.[11-21] It is presumed that the mode of binding of these agents involves partial intercalation of the acridine moiety between adjacent base pairs of DNA, with the aniline ring positioned in the minor groove of the helix.[21] This can easily be envisioned by examining the structure of the parent compound, 9-anilinoacridine (Figure 1).

Many qualitative structure–activity principles have emerged from the research, and most of these have been substantiated by quantitative SAR analysis.[16,21] Toxic and effective dosages of anilinoacridines have been shown to correlate well with electronic, steric, and lipophilic substituent parameters. In vivo studies have generally revealed a significant correlation between toxicity and antitumor potency. However, a common measure of tumor cell selectivity, ILS_{max} (maximal percent increase in life span at a fixed level of toxicity), is usually found to be independent of toxicity.[18] The aim of QSAR research on

Chart I. Substructures for Selected Environment Descriptors[a]



[a] An X indicates an unspecified atom type, and starred atoms are not checked for substituents during substructure searching.

antitumor compounds should ideally be directed at the development of more selective rather than more potent (and thus more toxic) derivatives. Consequently, the prediction of selectivity is an important objective of such research. Many attempts to derive quantitative relationships involving ILS_{max}, although sometimes satisfactory for small groups of homologous structures,[16] have resulted in equations with explained variances (R^2) of only 0.7 or lower, when a variety of types of compounds and substituents have been considered.[18]

In such cases, inferences may still be drawn from significant regression coefficients, but the relatively poor correlations are of little predictive value. It would be desirable to have some relationship that could, at the very least, separate active from inactive compounds (as indicated by ILS_{max}). Pattern-recognition techniques are ideally suited to this purpose, and in the past they have been used successfully to solve QSAR problems that could not be handled by conventional regression analysis methods.[22] We felt that the ADAPT system,[23] combining the capabilities of structure entry, molecular descriptor generation, and pattern-recognition analysis, might prove more successful in the study of the 9-anilinoacridines. Accordingly, we investigated a number of these compounds to try to relate tumor selectivity to molecular structure. This paper presents a summary of the results that were obtained.

Experimental Section

Methodology. A list of 776 9-anilinoacridines was compiled, containing toxicity (LD_{10}) and tumor selectivity (ILS_{max} at the LD_{10} dose) values.[21] These compounds represented a majority of the structures that have been evaluated by Cain, Denny, and co-workers, and a wide variety of substituents and structural variants were present. A histogram of the ILS_{max} values showed a large cluster of compounds at 25%, and the next major cluster appeared at 50%. A value of 35%, near the average of these, was

(7) P. Lettre, *Z. Physiol. Chem.*, **271**, 192 (1941).
(8) H. Blumenthal, *Bruns' Beitr. Klin. Chir.*, **154**, 50 (1931).
(9) Bruce F. Cain, Graham J. Atwell, and R. N. Seelye, *J. Med. Chem.*, **14**, 311 (1971).
(10) Bruce F. Cain, Graham J. Atwell, and R. N. Seelye, *J. Med. Chem.*, **12**, 199 (1969).
(11) Graham J. Atwell, Bruce F. Cain, and R. N. Seelye, *J. Med. Chem.*, **15**, 611 (1972).
(12) Bruce F. Cain, R. N. Seelye, and Graham J. Atwell, *J. Med. Chem.*, **17**, 922 (1974).
(13) Bruce F. Cain, Graham J. Atwell, and William A. Denny, *J. Med. Chem.*, **18**, 1110 (1975).
(14) Bruce F. Cain, Graham J. Atwell, and William A. Denny, *J. Med. Chem.*, **19**, 772 (1976).
(15) Graham J. Atwell, Bruce F. Cain, and William A. Denny, *J. Med. Chem.*, **20**, 1128 (1977).
(16) William A. Denny and Bruce F. Cain, *J. Med. Chem.*, **21**, 430 (1978).
(17) Bruce F. Cain, Bruce C. Baguley, and William A. Denny, *J. Med. Chem.*, **21**, 658 (1978).
(18) William A. Denny, Graham J. Atwell, and Bruce F. Cain, *J. Med. Chem.*, **22**, 1453 (1979).
(19) Bruce C. Baguley, William A. Denny, Graham J. Atwell, and Bruce F. Cain, *J. Med. Chem.*, **24**, 170 (1981).
(20) Bruce C. Baguley, William A. Denny, Graham J. Atwell, and Bruce F. Cain, *J. Med. Chem.*, **24**, 520 (1981).
(21) William A. Denny, Bruce F. Cain, Graham J. Atwell, Corwin Hansch, Augustine Panthananickal, and A. Leo, *J. Med. Chem.*, **25**, 276 (1982).
(22) Bruce R. Kowalski, Ed., "Chemometrics: Theory and Application", American Chemical Society, Washington, DC 1977; *ACS Symp. Ser.*, **no. 52** (1977).
(23) Andrew J. Stuper, William E. Brugger, and Peter C. Jurs, "Computer Assisted Studies of Chemical Structure and Biological Function", Wiley-Interscience, New York, 1979.

chosen as the cutoff for inactive compounds. In the entire list, there were 555 tumor-active compounds (ILS_{max} greater than or equal to 35%) and 221 inactives. Using a random number generator, a selection of 220 compounds gave a worklist of 158 active and 62 inactive compounds. Seven 10-methylacridinium compounds (five active, two inactive) were deleted from the sample, since some of the ADAPT programs do not support formally charged structures. The final training set of compounds thus contained 153 active and 60 inactive compounds. The structures of these compounds, with corresponding ILS_{max} values, appear in Table I.

The structures were entered into the ADAPT system, and four classes of descriptors were generated for use: (1) fragment descriptors, (2) topological descriptors, (3) substructure environment descriptors, and (4) physicochemical property descriptors (Table II). A fifth class of descriptors, geometric in nature, and derived from three-dimensional structures of modeled compounds, can also be generated in ADAPT. The size and flexibility of the many substituents in the present data set made it difficult to obtain valid models by automatic calculation, so geometric descriptors were omitted from the present study.

The fragment descriptors, which can be generated automatically, include various atom, bond, and ring counts and molecular weight. The topological descriptors include all-path counts[24] and a select number of molecular connectivity values.[25] For environment descriptors, the structures of misclassified compounds were periodically examined during the analysis, to identify substructrual fragments that could be used to better classify the compounds (e.g., Chart I). The substructures were encoded using first-order molecular connectivity calculations, which consider both the given substructure and its immediate environment in the molecule.[23]

The physicochemical descriptors that were generated included molar refractivity,[26] Bondi molecular volume,[27] the del-Re σ electronic charges at various positions on the aniline and acridine rings,[28] and the calculated log P values of the molecules.[29] Initially, the log P descriptor showed a surprisingly low correlation with the $\sum\pi$ values of the compounds ($R^2 = 0.72$). When corrections were made for ionizable groups, this correlation became expectedly higher ($R^2 = 0.98$).

Before any given descriptor was entered into the pattern-recognition analysis, the number of nonzero values was checked to see that it exceeded 10% of the total number of compounds and that nonzero values were present in both the active and inactive classes. Also, any descriptor was eliminated from consideration if it showed a high multiple correlation with other descriptors already in the analysis (e.g., R^2 greater than around 0.95). Out of a total of 84 descriptors generated, only 49 were actually used in the analysis (Table II), and no more than 30 to 35 descriptors were ever under consideration at the same time. All descriptors were autoscaled to zero mean and unit standard deviation for use in the pattern-recognition analysis.

The variance method of feature selection,[30] combined with the linear learning machine, was the primary technique used to select and reduce the number of descriptors. At any given step in the analysis, either the iterative least-squares[31] or the adaptive least-squares[32] learning machine was used to find the largest linearly separable subset of compounds that could be correctly classified using the current set of descriptors. Then, using a holdout technique, between 15 and 30 randomly chosen training and prediction subsets of the linearly separable group of compounds were generated. The learning machine was applied to each of these training sets until separation was achieved, generating 15 to 30 weight vectors. A mean weight vector was calculated, and the variance of each coefficient was determined (cf. Table IV). The descriptor corresponding to the coefficient with the largest variance was deleted from the descriptor set, and the process was repeated. In this fashion, "noisy" descriptors, whose coefficients fluctuated from one training set to the next, were removed from the descriptor pool.

This holdout method was also used to check the internal consistency of the descriptor sets. The weight vectors generated from the various training sets were used to predict the class membership of the prediction set compounds that were withheld. Each compound was withheld into only one of the prediction sets, and pooling the prediction set results gave a relatively unbiased estimate of the rate of misclassification for the given set of descriptors.

Results and Discussion

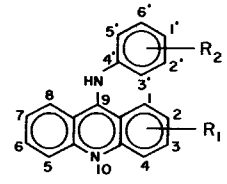
Using the methods described, a set of 18 descriptors was obtained which could correctly classify 200 (94%) of the compounds in the analysis. These descriptors appear in Table III, with substructures for the environment descriptors in Chart I. A principal components analysis on the correlation matrix of these descriptors showed the first 16 eigenvalues to be greater than 1.0, accounting for over 99% of the variance. The first two of these eigenvalues (values 2.9 and 2.3) accounted for only 58% of the total variance. In such a case, projection plots of the points in the space of the eigenvectors cannot adequately represent the compounds in relation to each other, so this plot was not included here. The several significant eigenvalues arise because of the presence of the uncorrelated environment descriptors. The implication of a large number of significant eigenvalues is that the descriptors effectively span the entire space represented by the compounds.[33]

Although they were selected solely for their ability to separate the active and inactive compounds, many of the descriptors in Table III can be explained in terms of what has been discovered about the mechanism of action of the 9-anilinoacridines. For example, the number of rings has been shown to be a determinant of the strength of binding to DNA when comparing 9-anilinoacridines with aminoacridine[12] and with 4-aminoquinolines.[17] Likewise, molar refraction, a measure of both volume and electronic polarizability, has recently appeared as a predictor variable in QSAR studies of the binding of derivatives of 4′-(9-acridinylamino)methanesulfon-*m*-anisidide (*m*-AMSA) to DNA.[19-21] The partial charges at the 2′-, 3′-, and especially the 6′-aniline positions have been implicated in the interaction of 9-anilinoacridines with polar groups in the minor groove of the DNA helix.[13] The σ charge at the 6′-position was not retained as a descriptor. However, descriptor 18 (σ charge at the aniline 2′-position) clearly reflects the influence of substituents at the 1′-, 2′-, and 3′-positions, which are the most commonly substituted ones. It is possible that the influence of 2′- and 3′-substituents on the σ charge at the 6′-position could not be detected by the relatively simple del-Re method of calculation that was used. The electronic descriptors on the acridine ring (positions 2 and 3) also correspond to the most highly substituted positions.

All the electronic descriptors were computed assuming a neutral charge on the molecules, which is probably not the case under physiological conditions.[15] It has been

(24) M. Randic, G. M. Brissey, R. B. Spencer, and C. L. Wilkins, *Comput. Chem.*, 3, 5 (1979).
(25) Lemont B. Kier and Lowell H. Hall, "Molecular Connectivity in Chemistry and Drug Research", Academic Press, New York, 1976.
(26) A. I. Vogel, "Textbook of Practical Organic Chemistry", Longman, New York, 1978, p 1035.
(27) A. Bondi, *J. Phys. Chem.*, 68, 441 (1964).
(28) G. del-Re, *J. Chem. Soc.*, 4031 (1958).
(29) J. T. Chou and Peter C. Jurs, *J. Chem. Inf. Comput. Sci.*, 3, 172 (1979).
(30) G. S. Zander, A. J. Stuper, and P. C. Jurs, *Anal. Chem.*, 47, 1085 (1975).
(31) Lucio Pietrantonio and Peter C. Jurs, *Pattern Recognition*, 4, 391 (1972).
(32) Ikuo Moriguchi, Katsuichiro Komatsu, and Yasuo Matsushita, *J. Med. Chem.*, 23, 20 (1980).
(33) Arthur Cammarata and Govind K. Menon, *J. Med. Chem.*, 19, 739 (1976).

Table I. Compounds in the Training Set



no.	R_1	R_2	ILS_{max}[a]	no.	R_1	R_2	ILS_{max}[a]
			Active Compounds				
1	3-C_2H_5	1'-CH_2CO_2H	200	78	3-$NHCH_3$	1'-$NHSO_2CH_3$-3' OCH_3	83
2		1'-$(CH_2)_5CO_2H$	200	79	4-$CONHCH_2CHOHCH_2OH$	1'-$NHSO_2CH_3$-3'-OCH_3	83
3	3-NO_2	1'-$NHCONH(p\text{-}C_6H_4$-guanidyl)	200	80	3,4-α-pyrido	1'-$NHSO_2CH_3$-3'-OCH_3	83
4	3-NO_2	1'-$NHSO_2(CH_2)_3$-guanidyl	200	81		1'-$(CH_2)_4CONH_2$	82
5	3-NO_2	1'-$NHSO_2(p\text{-}C_6H_4\text{-}NO_2)$	200	82	3-NO_2	1'-$NHCONH[p\text{-}C_6H_4\text{-}CH_2CH(NH_2)CO_2H]$	80
6		1'-$NHCONH[p\text{-}C_6H_4\text{-}N(CH_3)_2]$	170	83		1'-$NHCONH[p\text{-}C_6H_4\text{-}(CH_2)_4CO_2H]$	80
7		1'-$NHCONH[p\text{-}C_6H_4\text{-}(CH_2)_4NH_2]$	170	84	4-OC_4H_9	1'-$NHSO_2CH_3$-3'-OCH_3	80
8		1'-$NHCONH[p\text{-}C_6H_4\text{-}(CH_2)_5NH_2]$	170	85	4-OC_4H_9	1'-$NHSO_2C_3H_7$-3'-OCH_3	80
9	3-$NHCH_3$-5-CH_3	1'-$NHSO_2CH_3$	170	86	3-I-5-OCH_3	1'-$NHSO_2CH_3$	80
10	3-NO_2-6-CH_3	1'-$NHSO_2CH_3$	170	87	3-NH_2-4,5-$(CH_3)_2$	1'-$NHSO_2CH_3$-3'-OCH_3	80
11[b]	3-NO_2	1'-$NHSO_2(CH_2)_2NHCOCH_2NH_2$	170	88	3-NO_2	1'-$NHSO_2(CH_2)_5NH_2$	80
12	4-$CONH_2$	1'-$NHSO_2C_2H_5$-3'-OCH_3	166	89		1'-$NHSO_2C_4H_9$	78
13		1'-$(CH_2)_2CO_2H$	162	90	2-NH_2	1'-$NHCO(p\text{-}C_6H_4\text{-}NH_2)$	76
14	3-I	1'-$NHSO_2CH_3$	160	91	3-NO_2	1'-$NHSO_2CH_3$-3'-OCH_3	76
15	3-$NHCOCH_3$-6-NO_2	1'-$NHSO_2(CH_2)_3$-guanidyl	160	92	4-aza	1'-$NHSO_2C_5H_{11}$-3'-OCH_3	76
16		1'-$NHSO_2[p\text{-}C_6H_4\text{-}NHCOCH_3]$	151	93	4-CH_3	1'-$NHSO_2C_2H_5$-3'-OCH_3	76
17	3-NO_2-5-CH_3	1'-$N(COCH_3)SO_2CH_3$	150	94	3-OCH_3-5-CH_3	1'-$NHSO_2CH_3$	75
18		1'-$NHSO_2CH_3$-2'-aza	141	95		1'-$(CH_2)_7CO_2H$	75
19	3-$NHCOCH_3$-6-NO_2	1'-$NHSO_2CH_3$	141	96		1'-$(CH_2)_9CO_2H$	75
20	2-NH_2-3-Cl	1'-$NHSO_2CH_3$	140	97		1'-$NHSO_2CH_3$-3'-OC_2H_5	75
21	3-$NHCOCH_3$	1'-$NH(CO\text{-}i\text{-}C_3H_7)SO_2CH_3$	140	98	3-Cl-5-CH_3	1'-$NHSO_2CH_3$-3'-OCH_3	75
22	3-NH_2	1'-$NHSO_2CH_3$	134	99		1'-$NHSO_2C_6H_5$	75
23	3-NO_2-5-CH_3	1'-$NHSO_2CH_3$	130	100		1'-$(CH_2)_6CO_2H$	74
24	2,6-$(NH_2)_2$	1'-$NHSO_2(CH_2)_3$-guanidyl	129	101	3-NO_2	1'-$NHSO_2C_6H_{13}$-3'-OCH_3	74
25		1'-$NHSO_2(p\text{-}C_6H_4\text{-}SO_2CH_3)$	128	102	3-I	1'-$NHSO_2CH_3$-3'-OCH_3	74
26		1'-$NHCONH[p\text{-}C_6H_4\text{-}(CH_2)_2CO_2H]$	125	103[b]		1'-$CH{=}NNHCONH_2$	72
27	4-CH_3	1'-NH_2-3'-OCH_3	120	104	3-OH	1'-$NHSO_2CH_3$	72
28	4-NH_2	1'-$NHSO_2CH_3$-3'-CH_3	120	105	3-NO_2	1'-$NHSO_2CH_3$-2'-aza	72
29	3-NO_2-5-CH_3	1'-$NHSO_2CH_3$	120	106	3-NO_2	1'-$NHSO_2C_5H_{11}$-3'-OCH_3	72
30	2-NH_2-3-Br-5-CH_3	1'-$NHSO_2CH_3$	120	107	4-$CH_2N(CH_3)_2$	1'-$NHSO_2CH_3$-3'-OCH_3	72
31	3,4,5-$(CH_3)_3$	1'-$NHSO_2CH_3$-3'-OCH_3	120	108	3-NO_2-5-CH_3	1'-$NHSO_2CH_3$-3'-OCH_3	72
32	3,4-$(CH_3)_2$	1'-$NHSO_2CH_3$-3'-OCH_3	118	109	3-CH_3-5-OCH_3	1'-$NHSO_2CH_3$-3'-OCH_3	72
33	3-$NHCOCH_3$	1'-$NHSO_2C_2H_5$	115	110	4,5-$(CH_3)_2$	1'-$NHSO_2(CH_2)_4$-NH_2	72
34	3-$NHCOCH_3$	1'-$NHSO_2C_3H_7$	115	111	3-NO_2-5-CH_3	1'-$N(COC_2H_5)SO_2CH_3$-3'-OCH_3	72
35	3-NO_2-5-CH_3	1'-$NHSO_2CH_3$-3'-OCH_3	113	112	4-$CONHC_4H_9$	1'-$NHSO_2C_3H_7$-3'-OCH_3	71
36		1'-$NHSO_2CH_3$-3'-OCH_3	111	113		1'-$N(CH_2CH_2)_2NSO_2CH_3$	70
37		1'-$NHCOC_5H_{11}$	110	114		1'-$NHCOOCH_2CH(OH)CH_2OH$	70
38	3-NO_2	1'-$NHSO_2C_4H_9$	110	115	3-NO_2	1'-$NHSO_2CH_3$-3'-CH_3	70
39	4-OCH_3	1'-$NHSO_2CH_3$-3'-OCH_3	110	116	3-NO_2-4-CH_3	1'-$NHSO_2CH_3$	70
40	3-NH_2-5-CH_3	1'-$NHSO_2CH_3$	110	117	3-$NHCH_3$-4,5-$(CH_3)_2$	1'-$NHSO_2CH_3$	70
41	3-NO_2	1'-$NHSO_2(CH_2)_4$-guanidyl	110	118	3-NO_2	1'-$NHSO_2CH_3$	69
42	3-OCH_3	1'-$NHSO_2CH_3$	107	119		1'-NH_2-3'-OCH_3	66
43	4-$(CH_2)_2CONH_2$	1'-$NHSO_2CH_3$-3'-OCH_3	105	120		1'-$NHSO_2C_4H_9$	66
44	4-OCH_2CH_2OH	1'-$NHSO_2CH_3$-3'-OCH_3	105	121		1'-$NHSO_2C_6H_{13}$-3'-OCH_3	66

45	3-NH_2-6-OCH_3	1'-$NHSO_2CH_3$	105	122		1'-$CH{=}C(CN)_2$	65
46	2-NH_2-6-I	1'-$NHSO_2(CH_2)_2$-$NHCOCH_3$	105	123		1'-$N(CH_3)SO_2CH_3$	65
47		1'-$NHSO_2(CH_2)_2$-$NHCOCH_3$	105	124	4-OCH_3	1'-$NHSO_2C_4H_9$-3'-OCH_3	63
48	3-NH_2	1'-$NHSO_2CH_3$-3'-OCH_3	102	125		1'-$NHSO_2(CH_2)_2CON(CH_3)_2$	63
49	4-NH_2	1'-$NHSO_2CH_3$	100	126	3,6-$(NH_2)_2$	1'-SO_2NH_2	60
50	4-OC_3H_7	1'-$NHSO_2CH_3$	100	127		1'-$NHCOCH_3$	60
51	3-NH_2	1'-$N(CH_3)SO_2CH_3$	100	128		1'-NHCONH[p-C_6H_4-$(CH_2)_5CO_2H$]	60
52	4-$CONHCH_2CH_2OH$	1'-$NHSO_2CH_3$-3'-OCH_3	100	129	3-$NHCOCH_3$-6-Cl	1'-$NHSO_2CH_3$	60
53	3-NH_2-5-OCH_3	1'-$NHSO_2CH_3$	100	130	3-$NHCOCH_3$-4,5-$(CH_3)_2$	1'-$NHSO_2CH_3$	60
54	3-I-5-OCH_3	1'-$NHSO_2CH_3$-3'-OCH_3	100	131	3-NO_2-5,6-$(CH_3)_2$	1'-$NHSO_2CH_3$-3'-CH_3	60
55		1'-$NHSO_2$(p-C_6H_4-NH_2)	100	132	3,6-$(OCH_3)_2$	1'-$NHSO_2CH_3$	56
56	3-Cl	1'-$NHSO_2CH_3$-3'-OCH_3	99	133	3-NH_2	1'-SO_2NHCH_3	55
57	2-NH_2-3-Br	1'-$NHSO_2CH_3$-3'-OCH_3	99	134		1'-$NHSO_2CH_2Cl$	55
58		1'-CH_2SO_2NH(2-pyridyl)	98	135	4-CH_3	1'-$NHSO_2$(p-C_6H_4-NH_2)	55
59	3-$N{=}NN(CH_3)C_3H_7$	1'-$NHSO_2CH_3$	98	136	3-NH_2	1'-$N(COC_3H_7)SO_2CH_3$	55
60		1'-$(CH_2)_5CONH_2$	97	137	2-OCH_3-6-NO_2	1'-$NHSO_2CH_3$	55
61	4-$CONHC_4H_9$	1'-$NHSO_2CH_3$-3'-OCH_3	95	138	3-NH_2	1'-$COCH_3$	52
62	3-NO_2	1'-$SO_2(CH_2)_3NH_2$	95	139		1'-NHC_2H_5	52
63	3-$NHCOCH_3$	1'-$N(COC_3H_7)SO_2CH_3$	95	140	3-NH_2	1'-$SO_2NHC_5H_{11}$	50
64	3-CN	1'-$NHSO_2CH_3$-3'-OCH_3	93	141[b]		1'-OCH_3-2'-$NHSO_2CH_3$	50
65	4-$CONH_2$	1'-$NHSO_2C_5H_{11}$-3'-OCH_3	93	142		1'-$NHCON(CH_2CH_2)_2NCOOC_2H_5$	50
66	3-NO_2-6-CH_3	1'-$NHSO_2CH_3$-3'-OCH_3	92	143	2-NH_2	1'-$NHSO_2CH_3$	50
67		1'-CO_2H	90	144	4-$CONH_2$	1'-$NHSO_2CH_3$	50
68		1'-CH_3	90	145	3-Cl-6-OCH_3	1'-$NHSO_2CH_3$-3'-OCH_3	50
69		1'-$NHCOCH(C_2H_5)_2$	90	146	3-Cl	1'-$NHSO_2$(p-C_6H_4-NH_2)	50
70	2-NH_2-3-CF_3	1'-$NHSO_2CH_3$-3'-OCH_3	90	147	4-CH_3	1'-$N(COC_4H_9)SO_2CH_3$-3'-OCH_3	50
71	4-$CONH_2$	1'-$NHSO_2C_6H_{13}$-3'-OCH_3	87	148	1-NO_2	1'-$NHSO_2CH_3$-3'-OCH_3	50
72		1'-NHCONH(p-C_6H_4-CO_2H)	85	149	3-$NHCOCH_3$	1'-$NHSO_2C_6H_{13}$	42
73	3-CH_3	1'-$NHSO_2CH_3$	85	150		1'-$NHSO_2C_8H_{17}$-3'-OCH_3	40
74	4-OC_2H_5	1'-$NHSO_2CH_3$	85	151	3-i-C_3H_7	1'-$NHSO_2CH_3$-3'-OCH_3	40
75	2,6-$(NH_2)_2$	1'-$NHSO_2(CH_2)_2$-guanidyl	85	152	2,7-$(NH_2)_2$	1'-$NHSO_2CH_3$	35
76	3-CH_3	1'-$NHSO_2C_2H_5$	84	153[b]	2-OCH_3	1'-$NHSO_2CH_3$-3'-OCH_3	35
77	3-$NHCOCH_3$	1'-$NHSO_2C_4H_9$	83				

Inactive Compounds

154	3-NO_2	1'-OCH_3	30	184	2-$C(CH_3)_3$	1'-$NHSO_2CH_3$-3'-OCH_3	25
155		1'-$CH{=}CHCO_2H$-3'-OCH_3	27	185	2-$CONH_2$	1'-$NHSO_2CH_3$-3'-OCH_3	25
156		1'-$(CH_2)_8CONH_2$	26	186[b]	3-SO_2CH_3	1'-$NHSO_2C_6H_{13}$-3'-OCH_3	25
157		2'-$(CH_2)_2CO_2H$	25	187[b]	2-OCH_3-6-NH_2	1'-$NHSO_2CH_3$	25
158		2'-OH	25	188	2-OCH_3-6-$NHCOCH_3$	1'-$NHSO_2CH_3$	25
159		3'-NO_2	25	189	2-OCH_3-6-Cl	1'-$NHSO_2CH_3$-3'-OCH_3	25
160		3'-$NHSO_2CH_3$	25	190		1'-$NHSO_2CH(CH_3)CONH_2$	25
161		3'-$C(CH_3)_3$	25	191[b]		1'-$NHSO_2CH_2CH_2$(p-C_6H_4-NH_2)	25
162		3'-OCH_3	25	192[b]	4-$CONH_2$	1'-$NHSO_2(CH_2)_4NH_2$-3'-OCH_3	25
163		3'-F	25	193		1'-$NHSO_2(CH_2)_4$-guanidyl	25
164		1'-NH_2	25	194	3-Cl	1'-$NHSO_2(CH_2)_4$-guanidyl	25
165[b]		1'-NHC_4H_9	25	195		1'-$NHSO_2(CH_2)_2NHCO(CH_2)_2NH_2$	25
166[b]		1'-NHC_3H_7	25	196			25
167		1'-$N(C_2H_5)_2$	25	197		2'-OCH_3	25
168	3-NO_2	1'-$N(CH_2CH_2)_2NCH_3$	25	198		3'-C_2H_5	25
169[b]	3-NH_2	1'-NO_2	25	199		3'-OC_2H_5	25
170		1'-CH_2NH_2	25	200		1'-NH_2-2'-OCH_3	25
171		1'-$CH_2CH_2NH_2$	25	201[b]		1'-NHC_6H_{13}	25
172	3-NO_2	1'-$CH_2CH_2NH_2$	25	202		1'-$NHCOCH_2C_6H_5$	25
173	3-NO_2	1'-$(CH_2)_6NH_2$	25	203		1'-$CH_2N(CH_3)_2$	25

Table I (*Continued*)

no.	R_1	R_2	ILS_{max}[a]	no.	R_1	R_2	ILS_{max}[a]
174	3-$NHCH_3$	1'-CO_2H	25	**204**	3-NO_2	1'-$(CH_2)_4NH_2$	25
175		1'-OCH_3	25	**205**	3-NHCOCH_3	1'-$(CH_2)_2CO_2H$	25
176		1'-OCH_2CO_2H	25	**206**		1'-CH=CH-C_6H_5	25
177[b]	3-NH_2	1'-CO_2H	25	**207**		1'-F	25
178		1'-$CONHC_3H_7$	25	**208**		1'-$NHSO_2CH_3$-2'-Cl	25
179		1'-$NHSO_2CH_3$-3'-Cl	25	**209**	2-CH_3	1'-$NHSO_2CH_3$	25
180	1-NO_2	1'-$NHSO_2CH_3$	25	**210**	4-NO_2	1'-$NHSO_2CH_3$-2'-OCH_3	25
181	1-CH_3	1'-$NHSO_2CH_3$	25	**211**		1'-$NHSO_2(CH_2)_2NH_2$	25
182	3-$NHCOC_6H_5$	1'-$NHSO_2CH_3$	25	**212**		1'-$NHSO_2(CH_2)_5NH_2$	25
183	3-NHCOCH=CH-C_6H_5	1'-$NHSO_2CH_3$	25	**213**		1'-$NHSO_2(CH_2)_6NH_2$	25

[a] Maximal percent increase in life span at LD_{10} dosage level.[14] [b] Compound misclassified using the descriptors of Table III and the weight vector of Table IV.

Table II. Types of Descriptors Used in the Analyses

type	number of descriptors: generated	used[a]	final model
fragment[b]	18	9	2
substructure environment[c]	46	23	11
partial charge at ring position[d]	11	10	3
molecular connectivity	8	3	0
calculated log *P*[e]	1	1	0
molar refractivity	1	1	1
molecular volume[f]	1	1	0
molecular path counts[g]	2	1	1
total	84	49	18

[a] Some descriptors were removed prior to the pattern-recognition analysis, due to multiple correlations with other descriptors, or because of an insufficient number of nonzero values. [b] Including counts of various atom and bound types, rings, ring atoms, and molecular weight. [c] First-order molecular connectivity of a given substructure as imbedded in the total structure, including first neighboring atoms.[25] [d] Calculated by the del-Re method.[28] [e] Calculated by the fragment-additivity method of Leo.[29] [f] Bondi molecular volume at 0.75 × Van der Waals radius.[27] [g] See ref 24.

reported, however, in structure–activity studies on a series of diaminoquinolines that similar correlations were obtained, regardless of whether charged or uncharged molecules were assumed in the σ-charge calculations.[34] Presumably, the effects of substituents (which the σ-charge descriptors seem to be reflecting) would be similar on both charged and uncharged rings.[14]

The substructure environment descriptors in Table III show a progression from quite simple to relatively complex substructures. Whenever a given substructure was identified more than once in a particular molecule, the average value of the imbedded molecular connectivity for the substructure was stored as the descriptor value. Due to a limited number of possible values, some of the environment descriptors became simply indicator variables. This is reminiscent of Free–Wilson analysis, which, however, was not feasible for this data set, due to the large number of substituents and positions that were occupied.

Notable by its absence from Table III is the log *P* descriptor. Lipophilicity has been found to correlate with other measures of antitumor activity for the 9-anilinoacridines, when considering given structural types, or for homologous series of compounds. However, the relationship between activity and lipophilic character has often been lost or reduced when moving to a larger, more heterogeneous group of structures.[12,21] Thus, it is not surprising that lipophilicity did not remain in the descriptor set, considering the varied assortment of compounds in the analysis.

When the learning machine method was applied to the descriptors of Table III, it could correctly classify 149 of 153 (97.4%) active compounds and 51 of 60 (85%) inactive compounds. An average weight vector derived from hold-ten-out samples is shown in Table IV, along with measures of variation and relative importance of the coefficients. The structures that were misclassified are indicated in Table I. Eight of the compounds that were misclassified (compounds **165**, **166**, **169**, **177**, **187**, **191**, **192**, and **201**) have ionizable amino groups attached to either the aniline or acridine rings. These compounds formed

(34) George E. Bass, Donna R. Hudson, Jane E. Parker, and William P. Purcell, *J. Med. Chem.*, 14, 275 (1971).

Table III. Final Subset of Descriptors

no.	descriptor[a]	active class mean	active class SD	active class NNZ[b]	inactive class mean	inactive class SD	inactive class NNZ	highest R^2 value simple[c]	highest R^2 value multiple[d]	raw values (65)
1	no. of S atoms	0.804	0.415	122	0.417	0.530	24	0.745 (12)	0.797	1.0
2	no. of rings	4.137	0.345	153	4.100	0.303	60	0.612 (3)	0.898	4.0
3	av no. of paths per atom[e]	107.9	20.30	153	94.14	18.98	60	0.707 (4)	0.953	115.71
4	molar refractivity	120.5	12.53	153	109.2	15.62	60	0.707 (3)	0.925	135.66
5	environment SS 1[f]	0.361	0.529	49	0.410	0.545	22	0.373 (8)	0.596	1.142
6	environment SS 2	0.452	0.824	36	0.246	0.636	8	0.112 (7)	0.338	1.914
7	environment SS 3	0.800	1.405	38	0.591	1.263	11	0.236 (4)	0.663	3.386
8	environment SS 4	0.358	0.745	29	0.188	0.569	6	0.373 (5)	0.646	0.0
9	environment SS 5	0.688	1.126	45	0.413	0.854	12	0.616 (10)	0.706	1.559
10	environment SS 6	0.482	0.934	33	0.276	0.772	7	0.616 (9)	0.745	0.0
11	environment SS 7	2.209	2.416	70	1.529	2.268	19	0.712 (13)	0.857	0.0
12	environment SS 8	3.351	2.206	113	1.567	2.164	21	0.745 (1)	0.817	5.127
13	environment SS 9	2.136	2.906	54	1.532	2.680	15	0.712 (11)	0.870	0.0
14	environment SS 10	4.097	3.416	91	6.221	1.942	55	0.138 (12)	0.318	0.0
15	environment SS 11	1.687	3.168	34	2.637	3.627	21	0.247 (17)	0.572	0.0
16	charge position 2[g]	−0.016	0.034	153	−0.015	0.039	60	0.187 (8)	0.310	−0.031
17	charge position 3	0.028	0.059	153	−0.004	0.048	60	0.247 (15)	0.449	−0.029
18	charge position 2′	−0.019	0.027	153	−0.008	0.044	60	0.021 (15)	0.061	−0.008

[a] Overall F statistic for between-group separation (based on Hotelling's T^2) is 7.03 (18,194 df). [b] Number of nonzero values. [c] Highest R^2 with any other single descriptor (descriptor number in parentheses). [d] Multiple squared correlation with all other descriptors in table. [e] Total number of paths in molecule (of all lengths)/total number of atoms (non-hydrogen).[24] [f] Substructures for environment descriptors in Chart I. [g] Partial σ electronic charge calculated by del-Re method. See Table I for numbering convention.

Table IV. Average Weight Vector[a]

descriptor	coefficient	variance	rel contribution[b]
1	0.169	0.012	22.5
2	−0.172	0.002	14.8
3	0.326	0.004	29.1
4	−0.213	0.010	16.9
5	−0.141	0.011	19.2
6	0.033	0.044	9.4
7	0.090	0.029	16.9
8	0.457	0.003	23.5
9	0.083	0.027	16.0
10	−0.129	0.017	14.1
11	0.224	0.003	25.8
12	0.086	0.061	17.8
13	−0.323	0.002	23.5
14	−0.241	0.017	20.7
15	0.128	0.004	20.2
16	−0.142	0.016	12.7
17	0.153	0.001	19.2
18	−0.333	0.018	16.0
constant	0.361	0.003	

[a] Values represent the average of 20 weight vectors obtained using hold-ten-out samples drawn at random from the separable subset of 200 compounds (see text and Table I). Coefficients refer to autoscaled descriptors. [b] Overall percent misclassified when the given descriptor is removed from the weight vector.

a cluster in the histogram of the discriminant scores. Aside from the presence of these amino groups, no single structural feature appears more than once among the misclassified compounds. Since amino groups also appear in many of the correctly classified compounds, various substructure environment descriptors that were designed to account for the amines did not improve the classification results for the amino analogues.

Other workers have reported anomalous results for simple amine derivatives. In a recent paper relating frame-shift mutagenicity and DNA binding affinity of some AMSA derivatives, it was found that several unsubstituted and monosubstituted amino compounds lay outside a "mutagenic window" of DNA binding constant values, while still possessing antitumor activity.[35] It is possible that amine groups (because of their ability to be protonated) are capable of altering the stereochemistry of binding of the acridines to DNA and, thus, affect the biological activity. The use of pK_a values could conceivably correct for the amino compounds and improve the classification results. Although it is possible to predict the pK_a values of 9-anilinoacridines using σ substituent constants,[21] the quality of the predictions varies depending on the nature and location of the substituents. Because of this and because the emphasis in this study was on the use of computer-generated descriptors, pK_a values were not used in the analyses.

Throughout the analysis, there was a general tendency to misclassify a larger proportion of inactive compounds than active ones. Although this likely reflects to some extent the difference in class sizes, it may mean that the descriptor selection was not optimal for both classes. Alternatively, it may mean that the data constitute the "asymmetric" case of Wold and Dunn,[36] especially since quadratic classification was able to reverse the tendency somewhat (Table V).

Specific attempts at asymmetric classification did not achieve nearly the same overall classification success as the linear learning machine. For this reason, the learning machine methods were the primary classification techniques that were used. The descriptor set of Table III may reflect this fact. Strictly for comparison purposes, a number of alternative pattern-recognition methods were applied to the data set using the descriptors of Table III. The results are in seen in Table V. Analyses were performed on the full set of 213 compounds (set A in Table V) and on the linearly separable subset of 200 compounds (set B). Although the differences are not extreme, it is evident that those methods which emphasize class separation (learning machines, Bayes linear, and Bayes quadratic) give better classification for this set of descriptors than do the nonseparative methods (KNN, SIMCA). It

(35) Lynnette R. Ferguson and Bruce C. Baguley, *Mutat. Res.*, 82, 31 (1981).
(36) William J. Dunn III and Svante Wold, *J. Med. Chem.*, 23, 595 (1980).

Table V. Classification Results Using Various Pattern-Recognition Methods[a]

method	set[b]	% correctly classified: active	inactive	overall
linear learning machine[c]	A	97.4	85.0	93.9
	B	100.0	100.0	100.0
Bayes linear[d]	A	81.1	83.3	81.7
	B	89.9	90.2	90.0
Bayes quadratic[d]	A	84.9	91.7	86.9
	B	89.3	98.0	91.5
K nearest neighbor[e]	A	85.6	65.0	79.8
	B	91.3	70.6	86.0
SIMCA[f]				
F criterion[g]	A	79.1	66.7	75.6
	B	84.9	56.7	77.0
SD criterion[h]	A	88.2	56.7	79.3
	B	82.4	66.7	77.9

[a] Using the descriptors of Table III. [b] Set A = entire set of 213 compounds. Set B = linearly separable set of 200 compounds (Table I). [c] Including adaptive least-squares[32] and iterative least-squares learning machines. [d] Using equal prior probabilities. [e] Best overall results were obtained at 1 nearest neighbor. [f] Four-component models were used for each class. See ref 36. [g] Results based on smaller F ratio. [h] Results based on smaller standard deviation.

is entirely possible that the situation could be reversed if a different set of descriptors were chosen.

It is necessary if any classification study to validate the results. This can be accomplished by checking the internal consistency of the original-sample results using a holdout technique or by presenting a completely new prediction set to the weight vector. Both methods were followed in this research.

As described under Methodology of the Experimental Section, the internal consistency of the data was evaluated by applying a hold-ten-out method to the subset of 200 linearly separable compounds. An average predictive ability of 93% (95% active, 86% inactive) was achieved, which is comparable to the original-sample results.

To further test the validity of the descriptors and the weight vector, we randomly chose an untested prediction set of 50 compounds (35 active, 15 inactive) from among the structures not originally included in the analysis. The structures of the prediction set compounds are seen in Table VI. The descriptors in Table III were generated for these compounds, and the values were autoscaled using the means and standard deviations of the original training set data. Then, the average weight vector of Table IV was applied to the prediction data. Of the active compounds, 30 of 35 (86%) were correctly classified, while only 11 of 15 (73%) inactive compounds could be correctly placed. This is 82% correct overall, which well exceeds the 58% that would be expected using random selection corrected for class sizes.[37] When each of the individual weight vectors used to derive the average vector of Table IV was applied to the prediction compounds, essentially the same results were obtained (active, 24–26 compounds correctly classified; inactive, 11–12 correct).

A second prediction set was drawn from two recently published papers.[38,39] All the compounds reported in these articles that do not appear in Tables I or VI were included, generating a list of 69 structures (45 active, 24 inactive).[40] When the weight vector of Table IV was applied to the autoscaled descriptors for these compounds, 77% were correctly classified (78% active, 75% inactive). These results are lower than those of bothe the training set and the previous prediction set, but they can still be considered significant improvements over chance prediction (55% in this case). It is possible that the reduced prediction rates arise because the compounds in the second prediction set are not highly representative of the previous training and prediction sets. All the structures in reference 38 are monosubstituted aniline compounds, while only 31% of the training set structures are of this type. The structures of ref 39 are all monosubstituted acridine derivatives of m-AMSA, but only 8% of the training set structures are of this type. A multivariate test for differences between the descriptor values of the second prediction set and those of the combined training and first prediction sets gave a significant F statistic (Hotelling T^2, $F_{18,313} = 6.7$, $p < 0.001$).[41] Since the weight vector was derived for the training set structures, its use on descriptors whose means and variances differ much from those of the training set is analogous to extrapolating a regression line, and poorer prediction results might be anticipated.

Although the weight vector of Table IV is less than optimal for the second prediction set, the descriptors that were selected are still useful for separating these compounds. This was confirmed by retraining the weight vector using the 69 prediction structures alone. Only three of the compounds were misclassified (95% correct). Also, when the weight vector was trained using all 332 compounds in the analysis, an average of 91% correct was obtained (94% active, 83% inactive). Taken all together, the prediction results suggest that in practice one could expect to correctly classify 75–90% of compounds tested, using the given descriptors. Feature selection and weight vector development using more or larger training sets would likely improve prediction results still further. Nevertheless, even the results obtained here must be considered encouraging.

In this research, we chose to view tumor cell selectivity as an all or none response for the purpose of the study. Whether such a view of tumor selectivity is valid from a mechanistic standpoint was not important to the analysis, since our goal was the classification of the compounds and not the interpretation of the descriptors. The descriptors were selected on the basis of this viewpoint, and there is no particular reason why they should show any correlations with the quantitative ILS_{max} values.

This was tested by regressing the 18 descriptors on the log ILS_{max} values. A multiple R of only 0.58 ($F_{18,194} = 5.6$) was obtained. Examining one subset of the compounds, 92 methanesulfonanilides, we improved the R value to only 0.70 in a 10-descriptor model. No further improvements could be obtained when all 49 valid descriptors were examined using forward and backward stepwise regression analyses.

As a further test of a possible relationship with quantitative ILS_{max}, a three-class problem was created by placing the 26 most active compounds ($ILS_{max} > 120\%$, determined from the distribution of the ILS_{max} values) into a separate class. The inherently multiclass techniques,

(37) A purely random choice would classify only 50% of the compounds correctly. Taking prior information about class sizes into account, one would expect to classify (35/50) × 35 + (15/50) × 15 = 29 of the compounds correctly, which is 58%.

(38) Bruce C. Baguley, William A. Denny, Graham J. Atwell, and Bruce F. Cain, *J. Med. Chem.*, **24**, 170 (1981).

(39) B. C. Baguley, W. A. Denny, and B. F. Cain, *J. Med. Chem.*, **24**, 520 (1981).

(40) From ref 38: compounds **3, 4, 6, 12, 13, 15, 18–25, 27, 29–35, 37–39, 41, 42, 44–46, 48, 49**; from ref 39: compounds **3, 4, 6, 12, 13, 15, 18–25, 27, 29–35, 37–39, 41, 42, 44–46, 48, 49.**

(41) Donald F. Morrison, "Multivariate Statistical Methods", McGraw-Hill, New York, 1976, pp 128–141.

Table VI. Prediction Set Compounds[a]

no.	R_1	R_2	ILS_{max}
		Active Compounds	
214	4-$OCH_2CHOHCH_2OH$	1'-CO_2H	170
215		1'-α-*d*-glucopyranosyl	60
216		1'-$(CH_2)_4CO_2H$	187
217		1'-CH=CH(*p*-C_6H_4-CO_2H)	158
218		1'-NH_2-3'-CH_3	76
219[b]		2'-NH_2	50
220[b]		1'-N(CH_3)$COCH_3$	50
221[b]		1'-$NHCOCH_2CH(CH_3)_2$	105
222	3-NH_2	1'-$NHCOC_6H_5$	115
223		1'-NH-$NHCOOC_2H_5$	75
224[b]		1'-NHCONH(*p*-C_6H_4-$CH_2CH_2NH_2$)	170
225	3-$NHCOCH_3$	1'-N($COOCH_3$)SO_2CH_3	125
226		1'-$NHSO_2CH_3$-3'-NH_2	105
227		1'-$NHSO_2CH_3$-3'-OH	85
228		1'-$NHSO_2CH_3$-3'-OCH_2CH_2OH	67
229	3-N=$NN(C_2H_5)_2$	1'-$NHSO_2CH_3$	95
230	3-CH_3	1'-$NHSO_2CH_3$-3'-OCH_3	168
231	3-C_2H_5	1'-$NHSO_2CH_3$-3'-OCH_3	123
232	3-Cl	1'-$NHSO_2C_2H_5$-3'-OCH_3	95
233	4-Br	1'-$NHSO_2CH_3$-3'-OCH_3	100
234	4-$CONH_2$	1'-$NHSO_2CH_3$-3'-CH_3	68
235	4-$CONHC_4H_9$	1'-$NHSO_2C_5H_{11}$-3'-OCH_3	45
236	4-$CON(CH_3)_2$	1'-$NHSO_2CH_3$-3'-OCH_3	123
237	3,4-$(CH_3)_2$	1'-$NHSO_2CH_3$	55
238	3,5-$(CH_3)_2$	1'-$NHSO_2CH_3$	70
239[b]	3-$NHCOCH_3$	1'-$NHSO_2CH_3$	60
240	3-NO_2-6-OCH_3	1'-$NHSO_2CH_3$	60
241	4,5-$(CH_3)_2$	1'-$NHSO_2C_2H_5$-3'-OCH_3	70
242	3-NH_2-5,6-$(CH_3)_2$	1'-$NHSO_2CH_3$	110
243	3-NO_2-4,5-$(CH_3)_2$	1'-$NHSO_2CH_3$-3'-OCH_3	56
244	4-CH_3	1'-$NHSO_2C_2H_5$	65
245		1'-$NHSO_2$(*p*-C_6H_4-CO_2H)	135
246	3-$NHCOCH_3$	1'-N(COC_6H_{13})SO_2CH_3	100
247	4-CH_3	1'-N(COC_3H_7)SO_2CH_3	75
248	3-NO_2-5-CH_3	1'-N(COC_3H_7)SO_2CH_3	85
		Inactive Compounds	
249		3'-N(CH_3)SO_2CH_3	25
250		3'-CH_2CO_2H	25
251		3'-CH=$CHCO_2H$	25
252[b]	3-NO_2	1'-NH_2	30
253[b]		1'-NH_2-3'-OCH_3	25
254		1'-piperazino	25
255[b]	3-NO_2	1'-CH_2NH_2	25
256		1'-$(CH_2)_6NH_2$	25
257[b]	4-CH_3	1'-N($COCH_2CH_2COOC_3H_7$)SO_2CH_3-3'-OCH_3	25
258		3'-I	25
259		1'-$OCH(CH_3)CO_2H$	25
260[b]	3-$NHCH_3$	1'-$O(CH_2)_3CO_2H$	25
261		3'-$NHCOCH_3$	25
262		3'-SO_2NH_2	25
263		1'-CO_2H-3'-CH_3	25

[a] See Table I for numbering conventions. [b] Misclassified using the descriptors of Table III and the weight vector of Table IV.

KNN, SIMCA, and adaptive least squares, were applied using the 18 descriptors of Table III. The overall classification results for this three-class problem were poorer than those obtained in the two-class analysis (KNN and SIMCA, 62–63% correct; ALS, 85% correct). However, the amount of overlap was found to be least between the inactive and the very active classes, as one might expect.

These results illustrate the difference between model-based approaches to SAR, exemplified by Hansch and Free–Wilson analyses, and the approach used in this research. The descriptors that are useful in one method may not be optimal for another, although some overlap could exist. In addition, the ILS_{max} values do not fulfill the usual requirements for QSAR analysis. They do not represent equimolar responses (they are instead equitoxic), and on a log scale they span a range of only 0.9 log unit. We feel that the more qualitative pattern-recognition approach that we followed here was a more suitable type of analysis for these data.

Summary and Conclusions

We have demonstrated the applicability of pattern-recognition techniques, combined with the automatic generation of molecular descriptors, to the study of a number of 9-anilinoacridine antitumor agents. A set of 18 descriptors was found that could separate inactive from active compounds. A number of simple amine derivatives were incorrectly classified as active compounds, and this indicated that perhaps pK_a values might be useful as a descriptor to correct for these agents. When applied to prediction sets of compounds, the discriminant was able to distinguish active from inactive structures with somewhat poorer classification results. Unlike correlation methods, which attempt to predict and to draw structural

or mechanistic inferences from the regression coefficients, pattern-recognition techniques are mainly predictive in nature. Their utility lies in the screening of proposed or existing agents for activity. In this respect, the tendency that was consistently observed in this study, to correctly predict a greater proportion of active compounds than inactive ones, could be a benefit. It is clearly better to misclassify in favor of active derivatives, to avoid missing potentially beneficial agents. This tendency was seen to extend to the prediction sets as well.

Recent studies with *m*-AMSA in phase II clinical trials have given both encouraging[42] and disappointing[43] results. Thus, the search for a clinically effective agent in this series is not yet complete. Based on the results that have been presented here, it is reasonable to expect that pattern-recognition techniques could play a useful role in the identification of active members in the future.

Acknowledgment. The PRIME-750 and MODCOMP-II computers, on which this research was performed, were purchased with partial financial support of the National Science Foundation. The work was also partially funded by the U.S. Environmental Protection Agency.

(42) S. S. Legho, G. R. Blumenschein, A. U. Buzdar, G. N. Hortobagyi, and G. P. Bodey, *Cancer Treat. Rep.*, **63**, 1961 (1979).

(43) David S. Carroll, Nancy Kemeny, Garrett Lynch, and Thomas Woodcock, *Cancer Chemother. Rep.*, **64**, 1149 (1980).

Structure–Activity Relationships in Potentially Hallucinogenic *N,N*-Dialkyltryptamines Substituted in the Benzene Moiety

Toni B. Kline,† Frederick Benington,* Richard D. Morin, and John M. Beaton

Neurosciences Program and Department of Psychiatry, School of Medicine, University of Alabama in Birmingham, Birmingham, Alabama 35294. *Received December 14, 1981*

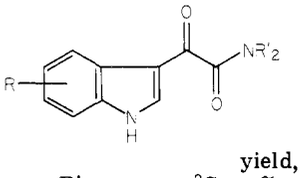
A series of *N,N*-dialkyltryptamines with methylthio or methylenedioxy substituents in the 4, 5, and 6 positions and methyl or isopropyl on the side-chain nitrogen has been synthesized. The behavioral pharmacology of these compounds showed them to possess Bovet-Gatti profiles characteristic of hallucinogens, and the 5-methylthio congener was the most potent. Binding studies at [^{3}H]LSD and [^{3}H]5-HT sites demonstrated that no single structural feature correlated with binding or behavioral changes and suggest a complex mode of action for these potential hallucinogenic agents.

Reports that tryptamines other than 5-hydroxytryptamine (5-HT) may be endogenous neuroregulatory agents,[1] complete with all appropriate biosynthetic and metabolic transformations,[2] have led us to prepare a series of *N,N*-dialkyltryptamines with novel substituents in the benzene moiety to be evaluated as hallucinogens. Recently reported methodologies in competitive binding studies[3] and measurement of behavior-disrupting activity in the rat make possible clearer distinction between behavioral and serotonergic effects of the compounds examined in this study.

Substituents, e.g., methoxy or hydroxy, in the 4, 5, or 6 positions of *N,N*-dimethyltryptamine induce significant changes in the neuropharmacological properties of these indolealkylamines.[4] In order to constrain *o*-methoxy groups into planar conformation, 4,5-(methylenedioxy)-*N,N*-dimethyltryptamine (1) and 5,6-(methylenedioxy)-*N,N*-dimethyltryptamine (2) were synthesized, and the pharmacological properties of 1 and 2 were compared with those of the known 4-methoxy- (3), 5-methoxy- (4), and 6-methoxy- (5) congeners.[5]

Significant changes in the potencies of substituted 2-phenylisopropylamines occurred when methylthio was substituted for methoxy.[6] Thioanisole partially exists in a rotated conformation in which the π system of the aromatic ring overlaps with the d orbitals rather than with the lone-pair p lobe, which has a rotational energy barrier of 2.05 kcal/mol,[7] slightly lower than most biological weak forces. It was therefore decided to synthesize the 4-, 5-, and 6-(methylthio)-*N,N*-dimethyltrytamines (6–8) in order to examine the differences between isosteres.

Table I. Ring-Substituted *N,N*-Dialkylindole-3-glyoxalamides



no.	R	R'	mp, °C	yield, %	formula[a]
22a	4,5-OCH_2O	CH_3	240–241	77	$C_{13}H_{12}N_2O_4$
22b	4,5-OCH_2O	i-C_3H_7	260 dec	56	$C_{17}H_{20}N_2O_4$
22c	5,6-OCH_2O	CH_3	217–220	79	$C_{13}H_{12}N_2O_4$
22d	5,6-OCH_2O	i-C_3H_7	278–280	81	$C_{17}H_{20}N_2O_4$
22e	4-SCH_3	CH_3	163–164	43	$C_{13}H_{14}N_2O_2S$
22f	4-SCH_3	i-C_3H_7	190–192	27	$C_{17}H_{22}N_2O_2S$

[a] IR and NMR spectra were consistent with structures given; the dried products were not analyzed but were reduced without further purification.

† Department of Chemistry, State University of New York at Stony Brook, Stony Brook, New York 11794.

(1) S. T. Christian, R. Harrison, E. Quayle, J. Pagel, and J. Monti, *Biochem. Med.*, 18, 164 (1977); S. A. Barker, J. Monti, and S. T. Christian, in "International Review of Neurobiology", J. R. Smythies and R. J. Bradley, Eds., Academic Press, New York, 1981, pp 83–110.

(2) S. A. Barker, J. A. Monti, and S. T. Christian, *Biochem. Pharmacol.*, **29**, 1049 (1980).

(3) S. Maayani, H. Weinstein, and J. P. Green, *Fed. Proc., Fed. Am. Soc. Exp. Biol.*, **38**, 376 (1979).

0022-2623/82/1825-0908$01.25/0