Computer Assisted Structure-Activity Studies of Chemical Carcinogens. An *N*-Nitroso Compound Data Set

J. T. Chou and Peter C. Jurs*

Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802.
Received February 9, 1979

N-nitroso compounds, consisting of nitrosamines and nitrosamides, are potentially important in the etiology of human cancer. An attempt to study the molecular structure–carcinogenicity relations of these compounds is reported. A pattern-recognition approach was used to develop predictive ability for carcinogenic potential. A set of 15 calculated molecular structure descriptors that supported a linear discriminant function able to successfully separate 116 carcinogens from 28 noncarcinogens was identified. Predictive ability of an overall of 91%—93% for carcinogens and 85% for noncarcinogens—was obtained in the randomized testing. This relatively high predictability demonstrates that pattern-recognition methods can be useful in analyzing these compounds for carcinogenic activity. The inclusion of two electronic descriptors implicitly supports the α-hydroxylation hypothesis. The relations of descriptors used and possible mechanism of action are discussed.

N-Nitroso compounds have been known since the preparation of dimethylnitroamine more than 100 years ago. The tumorigenic property of this compound was not described until 1956 by Magee and Barnes.[1] Since then, a large number of *N*-nitroso compounds have been synthesized and tested for carcinogenic activity by long-term administration of small doses to animals, mainly rats. This work has been done largely by Druckrey et al.[2] and Lijinsky and co-workers.[3] The carcinogenic activity of these compounds has been reviewed and discussed recently.[4]

The *N*-nitroso compounds, including nitrosamines and nitrosamides, are a class of potent and widespread environmental carcinogens and are potentially important in the etiology of human cancer.[5-7] They induce tumors in various vital organs of a wide range of animals.[4] The organotropic carcinogenic property was also noticed.[2] They have been found in betel nut,[8] in bacteria,[9] and in smoke and foodstuffs.[7] *N*-Nitroso compounds can be formed easily from the reaction of amines and nitrites.[10,11] The amine precursors are normal constituents of food, drugs, pesticides, and food additives, and nitrite is present abundantly in the environment, in cured meat, in human saliva, and from the reduction of nitrate.[7,8]

The mechanism of action for *N*-nitroso compounds at the molecular level has not been fully elucidated. These compounds may act primarily as alkylating agents to induce cancer,[12-14] although other mechanisms have also been proposed.[3,14] Metabolic activation is probably required for their carcinogenic activity.[2] Either α-hydroxylation or oxidative dealkylation may be the initial or the rate-limiting step toward nitrosamines' carcinogenesis.[2,15-21] The proposed proximate carcinogen, the highly reactive α-hydroxynitrosamine, may form and further decompose to a carbonium ion, the ultimate carcinogen, which could alkylate the critical cellular macromolecules to induce tumors genetically or epigenetically.[22,23] However, the oxidations in the β, ω, and ω-1 positions have been observed and may play some role as well.[24-27]

With the goal of understanding which molecular features could be responsible for carcinogenesis, the variation in activity from one compound to another, possibly the mechanism of action, and eventually predicting the carcinogenic activity of new compounds, several investigators have studied structure–activity relationships using a variety of approaches. The first attempt was made by Druckrey et al. who studied more than 60 *N*-nitroso compounds and proposed enzymatic α-hydroxylation to be the first step in the bioactivation of the compounds.[2] Recently, it has been demonstrated that the carcinogenicity of about 50 nitrosamines is inversely proportional to the number of carbon atoms of the alkyl chains.[28] Wishnok and co-workers have found a significant correlation between carcinogenic activity of nitrosamines and partition coefficients and the Taft polar substituent constants using a statistical approach.[29] Lijinsky and co-workers have been attempting to elucidate the chemical structure and carcinogenesis relations by systematically modifying the molecular structures of nitrosamines.[30,31]

In an attempt to further elucidate the carcinogenic potential of *N*-nitroso compounds and to predict the activity of new compounds, studies have been undertaken to study the molecular structure and carcinogenic activity relationships of *N*-nitroso compounds using pattern-recognition techniques. The bases and applications of pattern-recognition methodology to chemical and biological problems have been discussed and reviewed in a previous paper.[32]

Table I. Distribution of the *N*-Nitroso Compound Data Set

struct class	no. of compds[a]			
	+	w	−	?
dialkyl, diaryl, or mixed	47	6	18	3
piperidines	10	2	7	6
pyrrolidines	3	1	4	
morpholines	2	1	1	
piperazines	6		3	1
hydrazine, hydroxylamines	3		1	1
acyl alkyl	29	2	1	7
misc	5	1	0	
total	105	13	35	18

[a] Abbreviations used are: +, carcinogen; w, weak carcinogen; −, noncarcinogen; ?, controversial or have not been tested.

Experimental Section

The structure–activity studies on chemical carcinogensis of *N*-nitroso compounds reported in this paper were done using the ADAPT system,[33] which was run on the Department of Chemistry MODCOMP II/25 16-bit digital computer. The fundamental steps involved in performing an SAR study using this system are shown in Figure 1. The individual steps have been explained in detail previously.[32]

Data Set. The data used in this study were taken from several sources. Most of them were obtained with the help of Dr. Lijinsky of the Frederick Cancer Center. The others were taken from a published compilation[4] or papers.[34,35] A total of 171 compounds were entered into the ADAPT disk files. Entry of the structures was done by sketching them on the screen of a graphic display terminal under the control of an interactive program. The structure files are stored permanently on disk for further pro-

0022-2623/79/1822-0792$01.00/0

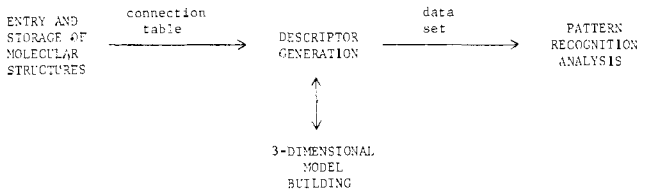
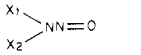
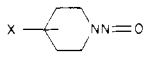
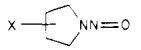
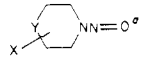
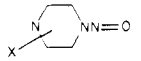
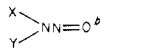
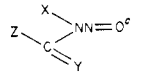
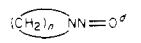


Figure 1. Flow chart of steps involved in structure–activity studies using chemical structure information handling and pattern-recognition methods.

Table II. Structures of *N*-Nitroso Compounds

compd	struct class
	dialkyl, diaryl, or mixed
	piperidines
	pyrrolidines
	morpholines
	piperazines
	hydrazine, hydroxylamines
	acyl, alkyl
	misc

[a] Y = O, S. [b] Y = O, N. [c] Y = O, N; Z = C, N, O. [d] $n > 5$.

cessing by other modules of the ADAPT system. The distribution and carcinogenic potency of the data set for each class of compounds are shown in Table I. The general structure for each class is shown in Table II. A subset of the total data set was used for study. It consists of 118 carcinogens (labeled + or w in Table I) and 35 noncarcinogens (labeled – in Table I). The identities of each of 153 compounds are listed in Table III.

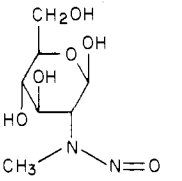
Descriptor Generator. After the data had been entered into the disk files correctly and a set of stored compounds had been chosen for investigation, the generation of molecular structure descriptors followed.

There are four general classes of descriptors (presently available in the ADAPT system): topological, geometrical, physicochemical, and electronic. These descriptors include (1) fragment descriptors, (2) substructure descriptors, (3) molecular connectivity descriptors, (4) environment descriptors, (5) geometric descriptors, and (6) σ-charge descriptors. The first five descriptors have been discussed in detail previously.[32]

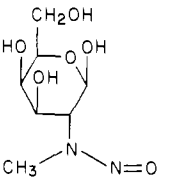
The σ-charge descriptor codes elementary electronic properties of the molecules. This descriptor's generation begins with a search of the structure being coded for the substructure of interest. If the substructure is absent, the descriptor value is zero. If the substructure is present, then the partial σ charge of each atom in the entire structure is calculated by the Del Re method.[36,37] At the user's option, the descriptor can be computed as (a) the average σ charge on all the substructure atoms as imbedded within the structure, (b) the charge of the substructure atom with the most positive σ charge, or (c) the σ charge on a specified substructure atom. In this study, only type a σ-charge descriptors are included in the final descriptor set.

For each of the 153 *N*-nitroso compounds many descriptors were developed and tested. After many trials of different subsets of the available descriptor set, a final set of 15 descriptors shown in Table IV was finally selected using the method described below. This set of descriptors provided the best overall performance of the pattern-recognition routines of any set of descriptors tested. Included are six fragments, five molecular connectivities (MC), and one geometric, one environment, and two σ-charge descriptors.

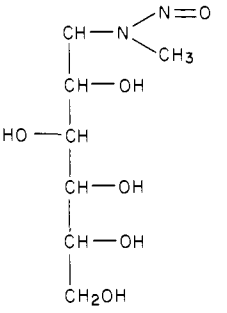
Chart I. Excluded Diastereomers



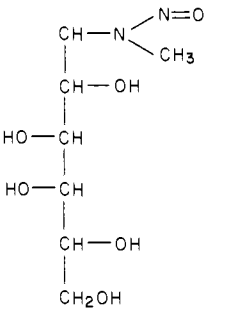
N-methyl-*N*-nitroso-β-D-glucosamine (carcinogen)



N-methyl-*N*-nitroso-β-D-galatosamine (noncarcinogen)

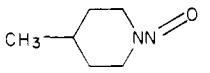


1-*N*-methyl-*N*-nitrosoamino-1-deoxy-D-glucitole (carcinogen)



1-*N*-methyl-*N*-nitrosoamino-1-deoxy-D-galatitole (noncarcinogen)

Table V gives some characteristics of the 15 descriptors. Mean values and standard deviations are given. Linear correlation coefficients were calculated for all pairs of descriptors, (15)(14)/2 = 105 pairs. The mean R value was 0.33. Seven of the correlation coefficients exceeded 0.85 for descriptor pairs that code similar information about the structures. These seven descriptor pairs and their R values are as follows: path 1 MC ring corrected with path 1 MC ring and heteroatom corrected (0.94); path 2 MC with path 3 MC (0.92); number of bonds with path 4 MC (0.90); number of carbons with path 1 MC ring and heteroatom corrected (0.88); number of carbons with number of bonds (0.87); path 1 MC ring and heteroatom connected with path 2 MC (0.87); number of bonds with path 1 MC ring corrected (0.85). Even with a few high linear correlation coefficients, however, each one of the 15 descriptors has been tested individually, and all 15 are necessary. The final column of Table V shows the descriptor values for one specific compound, 4-methyl-*N*-nitrosopiperidine.



Pattern-Recognition Analysis. Once each compound in the data set chosen for study was represented by a set of descriptors, then pattern-recognition methods were applied to analyze the data. The goal was to identify the common features among those coded in the data set and to develop discriminants which would separate the carcinogens from noncarcinogens. Various heuristic pattern-recognition methods were used. The detailed procedures for pattern-recognition analysis have been described previously.[32] The same procedures were used in this study.

There are two pairs of diastereoisomers, as shown in Chart I, in the data set. Since the descriptors currently in use cannot distinguish between diastereoisomers, these four compounds are excluded. Such a data set is called set A (149 compounds). A new training set with five misclassified compounds excluded from set A, called set B, is generated as a result of developing linear discriminants using an iterative least-squares algorithm. A complete linear separation of carcinogens from noncarcinogens was achieved when the linear learning machine was applied to set B. Set B then was used to determine which of the descriptors under investigation were worthy of retention using a linear learning machine algorithm. During the course of feature selection, a set of 15 descriptors was identified that was sufficient to separate 116 carcinogens from 28 noncarcinogens correctly. These are shown in Table IV.

Results obtained using different pattern-recognition techniques with training set A of 149 compounds and training set B of 144

Table III. Compounds Comprising the *N*-Nitroso Compound Data Set[a]

no.	compd	carcinogenicity	no.	compd	carcinogenicity
1	*N*-propyl-*N'*-nitro-*N*-nitrosoguanidine	+[b]	75	*N*-nitroso-3,6-dihydro-1,2-oxazine	+[c]
2	*N*-butyl-*N'*-nitro-*N*-nitrosoguanidine	+[b]	76	*N*-nitrosotetrahydro-1,2-oxazine	+[c]
3	*N*-isobutyl-*N'*-nitro-*N*-nitrosoguanidine	+[b]	77	*N*-nitrosodecamethylenimine	+[c]
4	*N*-pentyl-*N'*-nitro-*N*-nitrosoguanidine	+[b]	78	*N*,4-dinitroso-*N*-methylaniline	+[c]
5	nitrosopiperidine	+	79	1-[4-(*N*-methyl-*N*-nitrosamino)benzylidene]indene	+[c]
6	3-methyl-*N*-nitrosopiperidine	+	80	4-[4-methyl(*N*-nitrosamino)styryl]quinoline	+[c]
7	4-methyl-*N*-nitrosopiperidine	+	81	*N*-methyl-*N*-nitrosourea	+[c]
8	3-hydroxy-*N*-nitrosopiperidine	+	82	*N*-ethyl-*N*-nitrosourea	+[c]
9	4-hydroxy-*N*-nitrosopiperidine	+	83	*N*-methyl-*N*-nitrosourethane	−[c]
10	4-keto-*N*-nitrosopiperidine	+	84	*N*-ethyl-*N*-nitrosourethane	+[c]
11	4-chloro-*N*-nitrosopiperidine	+	85	2-chloroethyl-*N*-nitrosourethane	−[c]
12	3,4-dichloro-*N*-nitrosopiperidine	+	86	*N*-methyl-*N*-nitrosoacetamide	+[c]
13	3,4-dibromo-*N*-nitrosopiperidine	+	87	1,3-dimethyl-*N*-nitrosourea	+[c]
14	*N*-nitroso-1,2,3,6-tetrahydropyridine	+	88	*N*-nitrosotrimethylurea	+[c]
15	nitrosopyrrolidine	+	89	*N*-*n*-butyl-*N*-nitrosourea	+[c]
16	3,4-dichloro-*N*-nitrosopyrrolidine	+	90	*N*-nitrosophenylurea	+[c]
17	*N*-nitrosonornicotine	+	91	streptozotocin	+[c]
18	*N*-nitrosomorpholine	−	92	hydrazodicarboxybis(methylnitrosamide)	+[c]
19	2,6-dimethyl-*N*-nitrosomorpholine	−	93	*N*-methyl-*N'*-nitro-*N*-nitrosoguanidine	+[c]
20	dinitrosopiperazine	−	94	*N*-ethyl-*N'*-nitro-*N*-nitrosoguanidine	+[c]
21	2-methyldinitrosopiperzine	+	95	*N*-nitroso-2-imidazolidone	+[c]
22	2,5-dimethyldinitrosopiperazine	+	96	*N*-methyl-*N*-nitrosobiuret	+[c]
23	2,6-dimethyldinitrosopiperzine	+	97	*N*-ethyl-*N*-nitrosobiuret	+[c]
24	dinitrosohomopiperazine	+	98	*N*-methyl-*N*-nitroso-*N'*-acetylurea	+[c]
25	*N*-nitrosoazetidine	+	99	*N*-nitroso-*N*-methyl-*N'*-(2-benzothiazolyl)urea	+[c]
26	*N*-nitrosoheptamethylenimine	+	100	2-hydroxyethylnitrosourea	+[d]
27	*N*-nitrosohexamethylenimine	+	101	1-nitrosohydantoin	+[d]
28	*N*-nitrosooctamethylenimine	+	102	*N*-nitroso-5,6-dihydrouracil	+[d]
29	*N*-nitrosodimethylamine	+	103	1-naphthyl *N*-methyl-*N*-nitrosocarbamate	+[d]
30	*N*-nitrosodiethylamine	+	104	*N*-methyl-*N*-nitrosobenzamide	+[d]
31	*N*-nitrosobis(2-methoxyethyl)amine	+	105	ethylnitrosocyanamide	+[d]
32	*N*-nitrosobis(2-ethoxyethyl)amine	+	106	2-methyl-*N*-nitrosopiperidine	w
33	*N*-nitrosodi-*n*-propylamine	+	107	3,5-dimethyl-*N*-nitrosopiperidine	w
34	*N*-nitrosobis(2-oxopropyl)amine	+	108	*N*-nitroso-3-pyrroline	w
35	*N*-nitroso-*N*-methylethylamine	+	109	*N*-nitrosothiomorpholine	w
36	*N*-nitroso-*N*-methylundecylamine	+	110	nitrosododecamethylenimine	w
37	*N*-nitroso-*N*-methyldodecylamine	+	111	*N*-nitrosobis(2-chloroethyl)amine	w
38	*N*-nitroso-*N*-methyl-2-phenylethylamine	+	112	*N*-nitrosodiisopropylamine	w
39	*N*-nitroso-*N*-methylneopentylamine	−	113	*N*-nitrosodiisobutylamine	w
40	*N*-nitroso-*N*-methylphenylamine	−	114	*N*-nitrosobis(2-hydroxypropyl)amine	w
41	*N*-nitroso-*N*-methylcyclohexylamine	+	115	*N*-nitroso-*N*-ethylisopropylamine	w[c]
42	*N*-nitrosodi-*n*-butylamine	+[c]	116	*N*-nitrosodiethanolamine	w[c]
43	*N*-nitrosodi-*n*-pentylamine	+[c]	117	*N*,*N'*-dinitroso-*N*,*N'*-dimethylphthalamide	w[c]
44	*N*-nitroso-*N*-methylvinylamine	+[c]	118	*N*-nitrosohydantoic acid	w[d]
45	*N*-nitroso-*N*-methylallylamine	−[c]	119	2,6-dimethyl-*N*-nitrosopiperidine	−
46	*N*-nitroso-*N*-methyl-*n*-propylamine	+[c]	120	2,2,6,6-tetramethyl-*N*-nitrosopiperidine	−
47	*N*-nitroso-*N*-methyl-*n*-butylamine	+[c]	121	4-*tert*-butyl-*N*-nitrosopiperidine	−
48	*N*-nitroso-*N*-methyl-*n*-pentylamine	+[c]	122	2-carboxy-*N*-nitrosopiperidine	−
49	*N*-nitroso-*N*-(2-hydroxypropyl)-*n*-propylamine	+[c]	123	4-carboxy-*N*-nitrosopiperidine	−
50	*N*-nitroso-*N*-(2-oxopropyl)-*n*-propylamine	+[c]	124	nitrosoguvacoline	−
51	*N*-nitroso-*N*-methylbenzylamine	+[c]	125	methyl *N*-nitrosophenidylate	−
52	*N*,*N'*-dinitroso-*N*,*N'*-dimethylethylenediamine	+[c]	126	2,5-dimethyl-*N*-nitrosopyrrolidine	−
53	*N*,*N'*-dinitroso-*N*,*N'*-dimethyl-1,3-propanediamine	+[c]	127	2-carboxy-*N*-nitrosopyrrolidine	−
54	*N*-nitroso-*N*-ethylvinylamine	+[c]	128	2-carboxy-4-hydroxy-*N*-nitrosopyrrolidine	−
55	*N*,*N'*-dinitroso-*N*,*N'*-diethylethylenediamine	+[c]	129	*N*-nitrosophenmetrazine	−
56	*N*-nitroso-*n*-butyl-*n*-pentylamine	+[c]	130	2,3,5,6-tetramethyl-*N*,*N*-dinitrosopiperzine	−
57	*N*-nitroso-*N*-ethyl-*n*-butylamine	−[c]	131	nitrosopiperazine	−
58	*N*-nitroso-*N*-methylaminosulfolane	+[c]	132	4-methyl-*N*-nitrosopiperazine	−
59	*N*-nitroso-*N*,*N'*,*N'*-trimethylhydrazine	+[c]	133	*N*-nitrosobis(2-cyanoethyl)amine	−
60	*N*-nitroso-*N*-ethyl(2-hydroxyethyl)amine	+[c]	134	*N*-nitrosobis(2,2-diethoxyethyl)amine	−
61	*N*-nitrosobis(acetoxyethyl)amine	+[c]	135	*N*-nitroso-*N*,*O*-dimethylhydroxylamine	−
62	*N*-nitroso-*N*-*n*-butyl(4-hydroxybutyl)amine	+[c]	136	*N*-nitrosodi-*n*-octylamine	−
63	*N*-nitroso-*N*-*n*-butyl(3-carboxypropyl)amine	+[c]	137	*N*-nitrosodiallylamine	−[c]
64	*N*-nitroso-*N*-*n*-propyl(4-hydroxybutyl)amine	+[c]	138	*N*-nitrosodicyclohexylamine	−[c]
65	*N*-nitroso-*N*-(2-hydroxyethyl)-*n*-butylamine	−[c]	139	*N*-nitrosodiphenylamine	−[c]
66	*N*-nitroso-*N*-(2-oxopropyl)-*n*-butylamine	−[c]	140	*N*-nitrosodibenzylamine	−[c]
67	*N*-nitroso-*N*-methyl(2-chloroethyl)amine	+[c]	141	*N*-methyl-*n*-heptylamine	−[c]
68	*N*-nitroso-*N*-methylaminoacetonitrile	+[c]	142	*N*-nitroso-4-(methylamino)azobenzene	−[c]
69	*N*-nitrososarcosine	+[c]	143	*N*-nitroso-*N*-ethyl-*tert*-butylamine	−[c]
70	*N*-nitrososarcosine ethyl ester	+[c]	144	*N*-nitroso-*N*-*n*-butyl(3-hydroxypropyl)amine	−[c]
71	*N*-nitroso-*N*-ethyl-4-picolylamine	+[c]	145	*N*-nitrosodiacetonitrile	−[c]
72	*N*-methyl-*N*-nitroso-β-D-glucosamine	+[c]	146	*N*-nitroso-*N*-(1,1-dimethyl-3-oxobutyl)methylamine	−[c]
73	1-(*N*-methyl-*N*-nitrosoamino)-1-deoxy-D-glucitol	+[c]	147	4-(*N*-nitroso-*N*-methylamino)benzaldehyde	−[c]
74	*N*-nitroso-*N'*-carbethoxypiperazine	+[c]	148	*N*-methyl-*N*-nitroso-β-D-galactosamine	−[c]

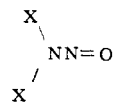
Table III (*Continued*)

no.	compd	carcinogenicity	no.	compd	carcinogenicity
149	1-(*N*-methyl-*N*-nitrosoamino)-1-deoxy-D-galactitol	–[c]	151	trinitrosohexahydro-1,3,5-triazine	–[c]
150	*N*-nitroso-L-proline ethyl ester	–[c]	152	*N*-nitrosoindoline	–[c]
			153	toluene-*p*-sulfonyl-*N*-methylnitrosamide	–[c]

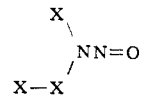
[a] Data are from Dr. Lijinsky, Frederick Cancer Research Center, unless specified otherwise. [b] Taken from ref 35. [c] Taken from ref 4. [d] Taken from ref 34.

Table IV. Descriptors Used to Represent the 153-Compound *N*-Nitroso Data Set

index	descriptors
1	no. of C atoms
2	no. of N atoms
3	no. of bonds
4	no. of dbl bonds
5	no. of aromatic bonds
6	no. of basis rings
7	path 1 mol connectiv (ring correct.)
8	path 1 mol connectiv (ring and heteroatom correct.)
9	path 2 mol connectiv
10	path 3 mol connectiv
11	path 4 mol connectiv
12	largest principle moment: intermed principle moment
13	environ of substruct



14	av σ charge of substruct



15	av σ charge of substruct

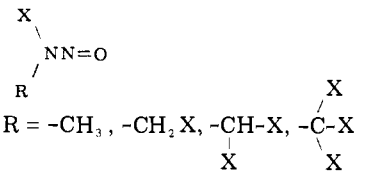


Table V. Mean and Standard Deviations for the 15 Descriptors and the Descriptor Values for 4-Methyl-*N*-nitrosopiperidine (MeNOPip)

descript no.	mean value	SD	value for MeNOPip
1	6.28	2.97	6
2	2.54	0.94	2
3	10.50	3.46	9
4	1.63	0.77	1
5	1.07	2.96	0
6	0.58	0.67	1
7	4.21	1.27	3.51
8	3.35	1.16	3.00
9	2.48	1.01	2.72
10	1.54	0.78	1.88
11	3.39	2.27	3.14
12	2.33	0.40	2.43
13	4.09	3.01	6.27
14	−0.032	0.02	−0.044
15	−0.047	0.05	−0.044

compounds are shown in Table VI. The linear learning machine and iterative least-squares algorithms achieve a perfect recognition rate on set B. All the other algorithms give less satisfactory results for both sets A and B. The *K*-nearest-neighbor method produced the least successful recognition rate.

Table VI. Pattern Recognition Results for the 153-Compound *N*-Nitroso Data Set Using 15 Descriptors

classifier	train. set used	no. correct. classif +	no. correct. classif −	% correct. classif +	% correct. classif −	% correct. classif total
Bayes (quadratic)	set A	109	26	94.0	78.8	90.0
	set B	110	25	94.8	89.3	93.8
Bayes (linear)	set A	108	26	93.1	78.8	89.9
	set B	107	24	92.2	85.7	91.0
K nearest neighbor						
K = 3	set A	112	11	96.6	33.3	82.6
K = 3	set B	113	11	97.4	39.3	86.1
iterative least squares	set A	115	27	99.1	81.8	92.8
	set B	116	28	100.0	100.0	100.0
simplex	set A	114	23	98.3	69.7	92.0
	set B	115	21	99.1	75.0	94.4
linear learning machine	set B	117	28	100.0	100.0	100.0

set A composition: no. in + class = 116; no. in − class = 33; total 149

set B composition: no. in + class = 116; no. in − class = 28; total 144

Table VII. Predictive Ability Tests for the 153-Compound *N*-Nitroso Data Set Using 15 Descriptors[a]

populations of training sets and prediction sets	carcinogens	noncarcinogens	total
training set	106	25	131
prediction set	10	3	13
total	116	28	144

predictive ability	no. predicted	no. correct	% correct
carcinogens	500	467	93.4
noncarcinogens	150	127	84.7
total	650	594	91.4

[a] Number of sets employed = 10; number of compounds excluded = 9.

After the feature selection process was completed and the set of 15 descriptors had been identified, the predictive ability of these descriptors was assessed. The results of this study are reported in Table VII. An overall rate of 91% was achieved. The predictive success for carcinogens was nearly perfect at 93%. However, the predictive ability for noncarcinogens was less successful at 85%. This skewed predictability for this data set indicates that when incorrect predictions are made the error is more likely to arise because a noncarcinogen is predicted to be a carcinogen than the reverse case. If errors are made, this is the preferable way to commit them.

The discriminant function developed by the linear learning machine which completely separates training set B could have been reported in this paper. However, we have not done so for

a number of reasons, including the following. First, the values of the discriminant function coefficients developed in a study are dependent on the set of compounds comprising the training set. Second, they are dependent on the details of implementation of the descriptor generation routines. Third, they would be somewhat dependent on the computer used for executing the routines due to round-off errors, etc. Therefore, the discriminant function developed in this work is specific to these conditions and would not have general applicability. To perform an SAR study of *N*-nitroso compounds or to predict the activity of a new *N*-nitroso compound would require access to much more than just the discriminant function coefficients developed in this work.

Discussion

The fact that a linear discriminant could be successfully developed to separate 116 carcinogenic *N*-nitroso compounds from 28 noncarcinogenic ones indicates that information concerning the carcinogenic potential is contained in the structures of these compounds. Although a chance correlation is possible in developing the discriminant functions, the ratio (about 10:1) of the number of compounds to descriptors per compound is such that such correlations were not responsible for the behavior of the discriminants.[38]

These structurally derived descriptors represent the composition of a structure. Descriptors 1–13 of Table IV also can be viewed as indicating steric, lipophilic, and electronic properties of a molecule, though descriptors 14 and 15 are electronic in nature. The simple fragmental descriptors, such as counts of atoms, bonds, and rings which are fundamental structural features of a molecule, are useful predictors in this study. The molecular connectivity indices have been shown to be highly correlated with a variety of biological activities,[39] including mutagenicity of nitrosamines.[40] It has been reported that the partition coefficients of nitrosamines are good descriptors for their carcinogenicity.[29,41] The molecular connectivity indices are also known to be well correlated with diverse physicochemical properties, such as partition coefficient.[39] The retention of molecular connectivity descriptors may show the same line of relations. The shapes of the molecules might be also important, since the geometric descriptor which represents the ratio of largest moment to the intermediate moment is retained.

N-Nitrosamides consisting of nitrosoureas, nitrosocarbamates, nitrosoguanidines and nitrosamides are proposed to be direct acting and not to require metabolic activation.[4,13] They decompose spontaneously in vivo and electrophilic alkylating species are thereby generated.[4,13] However, it has been suggested that some kind of bioactivation may be needed for their carcinogenesis.[3]

It is evident that α-hydroxylation of one of the alkyl chains of either dialkyl or cyclic derivatives of nitrosamines is very possibly the initial step or may be the essential step for their carcinogenesis and mutagenesis.[15,18,20,22] Nagata and Imamura have shown that the electronic structures of nitrosamines are important in determining the carcinogenic activity using molecular-orbital calculations.[42] The inductive effect of the substituent on the α carbon is also shown to be correlated with carcinogenic activity.[29] Most recently it has been demonstrated that metabolic hydroxylation of an aliphatic carbon atom showed a marked regioselectivity at the α carbon to heteratoms or π systems when quantum-mechanical approaches were used.[43,44] The retention of electronic descriptors (numbers 14 and 15 of Table IV) used in this study implicitly supports this α-hydroxylation hypothesis in view of the fact that these two descriptors represent the electronic properties of the N-NO group and the α-carbon atom.

After a carcinogen is administered, there are various events involved in the initiation of tumors, such as absorption, transport, distribution, metabolism, excretion, and interaction of the ultimate carcinogen with receptors. All these steps are important to different extents in determining the activity of a carcinogen. The importance of each contributing to the biological response varies from one compound to another. It is known that each of these events is very much dependent on the molecular structure and its physicochemical properties. If the manifestation of an in vivo biological response is not dominated by one or a few of these processes, all the events involved in causing the biological activity should be considered simultaneously in order to achieve a successful SAR study. In the case of carcinogenesis, it is even more complicated. In addition, the degree of repair of damaged DNA may be crucial[45] or promotion processes may also be required[23] in carcinogenesis. In view of the complexity and multivariate properties of carcinogenesis and unavailable quantitative data for most carcinogens, pattern-recognition methods should provide a very useful tool for a SAR study of carcinogens. In fact, the number of descriptors used in this study also may indicate the complexity of this problem.

Although the parameters used in the pattern-recognition study cannot be used to build an active or inactive molecule, they can be used to predict the activity of a hypothetical or new compound. The relatively high predictive ability (an overall rate of 91%) showed the usefulness of this approach.

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) P. N. Magee and J. M. Barnes, *Br. J. Cancer*, **10**, 114 (1956).
(2) H. Druckrey, R. Preussmann, S. Ivankovis, and D. Schmahl, *Z. Krebsforsch.*, **69**, 103 (1967).
(3) W. Lijinsky, *Prog. Nucleic Acid Res. Mol. Biol.*, **17**, 247 (1976).
(4) P. N. Magee, R. Montesano, and R. Preussmann, *ACS Monogr.*, **no. 173**, 491 (1976).
(5) W. Lijinsky and H. W. Taylor, *Cold Spring Habor Conf. Cell Proliferation*, **4**, 1579 (1977).
(6) P. N. Magee, *Food Cosmet. Toxicol.*, **9**, 207 (1971).
(7) R. Preussmann, *Arch. Toxicol.*, **suppl.** 1, 69 (1978).
(8) W. Lijinsky and H. W. Taylor, *J. Natl. Cancer Inst.*, **57**, 1315 (1976).
(9) A. D. Argoudelis, *J. Am. Chem. Soc.*, **89**, 4808 (1967).
(10) A. L. Fridman, F. M. Mukhametshin and S. S. Novikov, *Russ. Chem. Rev.*, **40**, 34 (1971).
(11) L. K. Keefer and P. P. Roller, *Science*, **181**, 1245 (1973).
(12) P. N. Magee and E. Farber, *Biochem. J.*, **83**, 114 (1962).
(13) P. N. Magee and J. M. Barnes, *Adv. Cancer Res.*, **10**, 163 (1967).
(14) E. K. Weisberger, *Annu. Rev. Pharmacol. Toxicol.*, **18**, 395 (1978).
(15) O. G. Fahmy, M. J. Fahmy, and M. Wiessler, *Biochem. Pharmacol.*, **24**, 2009 (1975).
(16) J. M. Rice, S. R. Joshi, P. P. Roller, and M. L. Wenk, *Proc. Am. Assoc. Cancer Res.*, **17**, 32 (1975).
(17) J. E. Baldwin, S. E. Branz, R. F. Gomez, and P. L. Kraft, *Tetrahedron Lett.*, 333 (1976).
(18) M. Wiessler and D. Schmahl, *Z. Krebsforsch.*, **88**, 25 (1976).
(19) J. Althoff, C. Grandjean, P. Pour, and B. Gold, *Z. Krebsforsch.*, **90**, 127 (1977).
(20) S. S. Hecht, C. B. Chen, and D. Hoffmann, *Cancer Res.*, **38**, 215 (1978).
(21) C. B. Chen, S. S. Hecht, and D. Hoffmann, *Cancer Res.*, **38**, 3639 (1978).
(22) A. M. Camus, M. Wiessler, C. Malaveille, and H. Bartsch, *Mutat. Res.*, **49**, 187 (1978).

(23) E. C. Miller, *Cancer Res.*, 38, 1479 (1978).
(24) J. Althoff, F. W. Krueger, and U. Mehr, *J. Natl. Cancer Inst.*, 51, 287 (1973).
(25) S. Dickhaus, G. Resznik, U. Green, and M. Ketkar, *Z. Krebsforsch.*, 91, 189 (1978).
(26) F. W. Kruger, *Z. Krebsforsch.*, 76, 145 (1971).
(27) M. E. Okada, T. Suzuki, J. Aoki, M. Iiyoshi, and T. Hashimoto, *GANN Monogr.*, no. 17, 161 (1975).
(28) J. S. Wishnok and M. C. Archer, *Br. J. Cancer*, 33, 307 (1976).
(29) J. S. Wishnok, M. C. Archer, A. S. Edelman, and W. M. Rand, *Chem.-Biol. Interact.*, 20, 43 (1978).
(30) W. Lijinsky, *New Sci.*, 27, 216 (1977).
(31) W. Lijinsky, in "Structural Correlates of Carcinogenesis and Mutagenesis", I. M. Asher and C. Zervos, Eds., The Office of Science, Food and Drug Administration, 1978, p 148.
(32) P. C. Jurs, J. T. Chou, and M. Yuan, *J. Med. Chem.*, 22, 476 (1979).
(33) A. J. Stuper, W. E. Brugger, and Peter C. Jurs, in "Chemometrics: Theory and Application", B. R. Kowalski, Ed., American Chemical Society, Washington, D.C., 1977.
(34) K. Lee, B. Gold, and S. S. Mirvish, *Mutat. Res.*, 48, 131 (1977).
(35) J. McCann, E. Choi, E. Yamasaki, and B. N. Ames, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 5135 (1975).
(36) G. Del Re, *J. Chem. Soc.*, 4031 (1958).
(37) A. J. Hopfinger, "Conformational Properties of Macromolecules", Academic Press, New York, 1973.
(38) A. J. Stuper and P. C. Jurs, *J. Chem. Inf. Comp. Sci.*, 16, 238 (1976).
(39) L. B. Kier and L. H. Hall, "Molecular Connectivity in Chemistry and Drug Research", Academic Press, New York, 1976.
(40) L. B. Kier, R. J. Simons, and L. H. Hall, *J. Pharm. Sci.*, 67, 725 (1978).
(41) G. M. Singer, H. W. Taylor, and W. Lijinsky, *Chem.-Biol. Interact.*, 19, 133 (1977).
(42) C. Nagata and A. Imamura, *Gann*, 61, 169 (1970).
(43) B. Testa, J. C. Bunzil, and W. P. Purcell, *J. Theor. Biol.*, 70, 339 (1978).
(44) B. Testa and D. Mihailova, *J. Med. Chem.*, 21, 683 (1978).
(45) J. L. Marx, *Science*, 200, 518, (1978).

Actinomycin D Oxazinones as Improved Antitumor Agents

Sisir K. Sengupta,*[1] Dorothy H. Trites, Maddula S. Madhavarao, and William R. Beltz

Division of Medicinal Chemistry and Pharmacology, Sidney Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, and Boston University Medical Center, School of Medicine, Boston, Massachusetts 02118.
Received August 18, 1978

1,4-Oxazinone derivatives of the phenoxazinone chromophore in actinomycin D (AMD) have been synthesized by condensation of AMD with α-keto acids. By varying the starting α-keto acid, the substitutions on the oxazinone ring and, consequently, the lipophilicity of the molecule could be altered. These oxazinone derivatives revert to AMD in physiological media and it appears that these oxazinones are "depot" forms of AMD and possess physicochemical and DNA-binding properties which are significantly different from those of AMD. The oxazinones, which have bulky and lipophilic substituents at position 3, demonstrate more pronounced antitumor activity against P388 mouse leukemia and are less toxic than AMD.

Actinomycin D (AMD, **1**) is one of a family of chromopeptide antibiotics isolated from *Streptomyces* cultures.[2-4] The naturally occurring actinomycin antibiotics have a common 2-aminophenoxazin-3-one ring system with two cyclic pentapeptide lactones attached at the 1 and 9 positions and differ from one another in one or two amino acids in the peptide lactones.[5-12] The integrity and configuration of the peptide rings, as well as the functions at the C_2, C_3, C_4, and C_6 positions on the phenoxazine rings, have been shown to be important for deoxyribonucleic acid binding[13-19] and the expression of their antibiotic activity.[13-19]

AMD has been used in the treatment of human neoplasia and is one of the few agents possessing curative effects against two different tumors:[20] Wilms' tumor[21] and gestational choriocarcinoma.[22] Although AMD is effective at remarkably low doses, its spectrum of antitumor activity is relatively narrow and its clinical administration is complicated by the high toxicity of the drug.[23]

For the past few years this laboratory has carried out investigations on AMD and 7-substituted AMD analogues.[24,25] These studies have established that 7-nitro-AMD and 7-amino-AMD, in spite of poor antibacterial properties, are indeed comparable to AMD in four transplantable mouse tumor systems. Another AMD analogue, N^2-(γ-hydroxypropyl)-AMD, exhibiting similar behavior, has been reported by Meienhofer and Johnson.[26] These findings confirm that the antibacterial activity of AMD analogues, long used to indicate biological usefulness of such compounds, does not provide reliable predictive data for other bioassay systems.

The 3-methyloxazinone of actinomycin C_2 was synthesized by Brockmann[27] and reported to be a poor antibacterial agent,[28] but in this laboratory it was found that a comparable analogue, **2b**, is similar to AMD in activity against AMD-sensitive P388 murine leukemia.

Our initial studies were done on actinomycin chromophore model compounds in which the pentapeptide lactones P in **1** (Scheme I) were replaced by $N(C_2H_5)$ groups.[25] On the basis of studies on these model derivatives,[29] the 3-methyloxazinone of AMD (**2b**) was synthesized (Scheme I), and it was observed that **2b** could be converted almost quantitatively to AMD in the presence of aqueous ethanol at pH 7.3 or above. In addition, it was observed that the 3-phenyloxazinone of AMD (**2g**) required a higher pH above pH 7.6 or a longer time for conversion to AMD in aqueous ethanol. These observations suggested that the oxazinones of AMD may act as depot or prodrug forms of AMD, provided they behave similarly in biological systems, and that the nature of the substituents at the 3 position of the oxazinones might influence their biological activity.

Synthesis and Physicochemical Properties. The successful synthetic route for **2a** and **2g** is shown in Scheme I; the results are summarized in Table I. AMD (**1**) was reduced catalytically under 1 atm of hydrogen pressure, and the reduced derivative was condensed with α-keto acids in methanol. Air was scrupulously excluded from the reaction atmosphere. The products are stable in acid and are usually purified on silicic acid columns. They are easily differentiated from the starting material, AMD, by TLC, UV, and NMR analysis. It was also noted that the specific rotation changes noticeably upon for-

0022-2623/79/1822-0797$01.00/0